The University of Reading

Met Office

# Numerical Weather Prediction

## The Summer 2003 Reruns with the High Resolution Trial Model

**Forecasting Research Technical Report No. 457**
**Joint Centre for Mesoscale Meteorology Report No. 151**

**H.W. Lean, S.P. Ballard, P.A. Clark M.A.G. Dixon, Zhihong Li and N.M. Roberts**

*email: nwp_publications@metoffice.gov.uk*

**©Crown Copyright**

**The Summer 2004 Reruns with the High Resolution Trial Model**

Forecasting Research Technical Report No. 457

Joint Centre for Mesoscale Meteorology Report No. 151

H.W.Lean, S.P.Ballard, P.A.Clark, M.A.Dixon, Zhihong Li and N.M.Roberts

Version History

| Version | Date | Approval | Changes |
|---|---|---|---|
| 1.0 | 1/04/04 | | Initial version |
| 2.0 | 1/07/04 | | Added sections on objective statistics |
| 3.0 | 5/11/04 | | Numerous changes to style and content. |
| 4.0 | 2/02/05 | P.A.Clark | Add CAPE figure and pmsl observational analysis. Description of assimilation rewritten. Further changes to style. |

Abstract

The High Resolution Trial Model (HRTM) suite consists of 12km, 4km and 1km models over the southern UK. This paper describes runs of these models for four cases from summer 2003 in order to assess the high resolution models performance in short range forecasting of convective events.. These reruns included comparison of a suite including continuous assimilation in the 4km and 1km models to a simpler suite where the high resolution models were spun up from the 12km analysis each cycle.

Both subjective and objective assessment of the forecasts imply that there are benefits from running at high resolution in particular for higher accumulation thresholds. The objective statistics reveal large differences between the four cases which implies that a much larger sample is required for reliable statistics. The spin up forecasts suffer from insufficient rain in the first few hours and correspondingly too much rain around the 5 hour forecast range. As would be hoped, the runs with assimilation don't suffer form these problems although they do have problems of their own with too much rain early in the forecasts. The encouraging results from these trials will be checked with more extensive trials on cases from summer 2004.

## 1. Introduction

With the introduction of the non-hydrostatic version (Cullen et al 1997) of the Unified Model (UM) and the continuing increase in available computer power the Met Office is looking to introduce models into its operational NWP suite of higher resolution than the current smallest gridlength of 12km. This paper reports some preliminary tests of a prototype forecasting system, the High Resolution Trial Model (HRTM), which consists of a 4km and a 1km gridlength model over the southern United Kingdom.

The HRTM project is a progression from the case-study based work on high resolution (24km up to 2km) versions of the UM which was carried out in JCMM from 2000 onwards (Lean and Clark 2003) This work was on a variety of cases and showed that high resolution models have the potential to improve forecasts and the representation of various phenomena. The HRTM project started in mid 2002 and aims to develop and evaluate 4km and 1km gridlength models in a scenario closer to an operational system. The emphasis of this project has been short range (out to T+6-9) forecasts of convective rain (particularly severe convective rain). The basic concept is to run the models on fixed domains for a large number of cases in order to gather verification statistics. These statistics are then used to evaluate the potential advantages of high resolution models. The first phase of the project consisted of a number of sensitivity studies in order to evaluate satisfactory values of a number of model parameters (timestep, horizontal diffusion etc). This paper reports work from the second phase of the project. In this phase data assimilation was added to the high resolution models and short trials were run of cases from the summer 2003 period.

## 2. Model Configurations

### (a) General

The 4km and 1km models were run in a very similar configuration to the (then) current operational 12km model. UM version 5.5 was used which is a non-hydrostatic version (Cullen et al 1997). The current operational versions of this model include a non-local first order boundary layer scheme (Lock et al 2000), a mass flux convection scheme with Convectively Available Potential Energy (CAPE) Closure (Gregory and Rowntree 1990) and a mixed phase precipitation scheme (Wilson and Ballard 1999). As discussed in section (c) below the 1km model was run without the convection scheme and the 4km with the convection scheme switched on but modified to produce less mass flux. The differences between the configuration of the 12km model and the 4 and 1km models are summarised in table 2.1. These parameters were arrived at by case study testing during the earlier phase of the project referred in section 1. The changes to the model domains are described in section (b) below and to the parameterisations in section (c).

In order to evaluate a possible operational system 6 hour forecasts were run at all three resolutions at 3 hour intervals over the period of interest in each case. The 12km model was rerun as well as the 4 and 1km ones in order to provide boundary conditions for the 4km domain and also to enable statistical comparison of the current model with the new high resolution ones. The 3 hour cycle length was chosen for convenience since that is the length currently used in the operational 12km model. It is recognised that a real operational system at high resolution may benefit from a shorter cycle length. A number of options were run including spinning up the 4km

and 1km forecasts every cycle from the 12km T+1 analysis and running continuous cycles at 4km and 1km. Section (d) discusses the assimilation options used in the models. Since several assimilation options were run for each case, in order to avoid duplicating runs, the boundary files and other output from the 12km and 4km runs were archived and retrieved when running inner models with different options. More details of the suites are given in appendix B.

(b) Domains

The domains used for the 4km and 1km models are shown in figure 2.1. The 4km model was one way nested inside the 12km UK mesoscale model. This domain was chosen so as to include most of the southern half of the UK and also a good deal of northern France and the southern approaches. Since the focus of the work was on severe convection this domain was chosen in order to encompass events of thunderstorms moving northwards from the south. This model was originally set up with 76 vertical levels (the 38 level set used in the 12km model doubled all the way up to the top of the model at 40km) but in the work reported here it was run with the 38 levels due to constraints of data assimilation. With 190x190x38 levels the 4km model is only about 50% larger than the UK mesoscale model in terms of 3D gridpoints (although the shorter timestep must also be taken into account when calculating CPU requirements).

The 1km model was one way nested inside the 4km model. The domain was chosen to be roughly centred on the Chilbolton radar and to cover part of southern England. At 300x300x76 levels it was around 5 times larger than the UK mesoscale model and was the largest model which was thought to be practical with the available computing resources.

For both the 4km and 1km models high resolution orography, land-sea mask and land use data were provided. Other fields were interpolated from the 12km data. For technical reasons the 1km model used orography data at only 2km resolution for the work presented here (more recently 1km resolution data has been incorporated). All the orography datasets were smoothed in an equivalent way to that used in the operational models.

(c) Parametrisations

In the 4km and 1km models changes have been made to the parameterisations for convection, diffusion/turbulence and microphysics. These are discussed in turn in this section.

The UM convection scheme was designed on the assumption that there are many cumulus clouds in each gridbox. This assumption is already dubious for the 12km model and is completely wrong for the 4km and 1km models. Since convection is the most important type of weather being forecast by these models it is important to address this issue.

For the 4km model it is found that running with the standard UM convection scheme (CAPE closure timescale set to 30min as in the 12km operational model) may give the correct general area of precipitation but, particularly for more severe convection, does not give sufficient intensity. In addition with the convection scheme the model can miss organisation of showers which may happen as they mature. This is a result of the

fact that the convection scheme works on each column in the model separately and does not take account of convection in nearby columns. Another problem is that in the 4km model the interaction between the convection scheme and model dynamics has been observed to generate spurious arcs of rain emanating from regions of convective activity.

Figure 2.2 shows the domain averaged rainfall rate as a function of time for a case of small, scattered showers which later on develop and become more organised (3rd May 2002). In this case the model with the convection scheme initially does well but the shower activity peaks at about the same time the solar heating peaks and then decays when in the radar data the rainfall rate continued to increase. This error is a manifestation of the fact that the convection scheme simply responds to the immediate forcing. It also fails to produce the increase in rain later in the day due to organisation of the showers as described above. As a result of this by 17 UTC the model averaged rain rate has fallen almost to zero whereas the radar was still seeing the rain rate increasing.

The obvious alternative to using the convection scheme is to run the model without it. This appears to give reasonable results when the showers are large and heavy but when the showers are weaker and more scattered they are either missed altogether or replaced by showers which are too large, too widely spaced and too heavy. This usually results in too much rain being produced overall in the model. The aliasing onto larger scales is exactly what would be expected from attempting to resolve showers which are too small compared to the gridlength of the model. Results for the model without a convection scheme for the scattered convection case are also shown in figure 2.2. The model in this case produces far too much rain (by up to a factor of two) in the later part of the simulation.

Figure 2.2 also illustrates the other major problem with explicit convection which is that the initiation of convection is delayed (in this case by over an hour). It is likely that this problem arises because, in the real atmosphere, convective initiation starts with perturbations on scales smaller than the model gridlength. Despite these problems the figure illustrates that running without a convection scheme does move the time of peak rainfall much closer to the correct time. It is generally observed that running without convection does produce improvements to forecasts in the sense that it can correct problems due to running with the convection scheme.

Since neither of the above options is entirely satisfactory an intermediate solution has been developed in which the convection scheme is run, but with a reduced intensity (Roberts 2003). The convection scheme uses a so-called ‘CAPE closure timescale’ ($\tau$) to determine the cloud-base mass flux (M) of the parameterized convective plume and therefore the intensity of the parameterized convection. The CAPE closure timescale $\tau$ is defined as the time it takes to remove all the Convective Available Potential Energy (CAPE) in a column – the shorter it is, the more active the convection.

$M \alpha CAPE / \tau$

The convection scheme in the 4-km model has been modified so that $\tau$ is a function of CAPE.

$\tau = t/c * CAPE + t * \exp(-CAPE/c)$ t and c are tuneable parameters.

The function is shown graphically in Figure 2.3. Instead of always becoming larger where the CAPE is larger, the cloud-base mass flux M is restricted to a limiting value.

The purpose of this approach is to allow the convection scheme to behave in the standard way when CAPE is small (similar to using constant $\tau$). In that situation, any convective clouds should be represented by the parameterization scheme because they are likely to be too small and weak to be resolved on the model grid. However, when the CAPE is larger, convection is more intense and may be explicitly resolved on the grid. In that case, the activity of the convection scheme is limited to allow more of the convection to be resolved rather than parameterized. The benefit of explicitly resolving the convection where instability is sufficiently large is that it is possible to model the life-cycle and organisation of convective clouds. This should lead to better forecasts of the more intense convective events.

Figure 2.2 gives an example of how the modified CAPE closure behaves in practice. Introducing the CAPE dependent CAPE closure scheme with a small value of c (as shown) can give correct amounts of rain later in the day but at the cost of making the delay in initiation even worse (half hour longer in this case). If c is increased the curve moves towards the standard convection scheme curve which, as noted above, is more realistic at the start (by the convection being handled by the convection scheme) but then the rainfall rate is too small later. The runs described in this paper were carried out with the intermediate scheme with parameters set to c=0.01J/Kg and t=1200s (20mins). The small value of c means that these runs are close to the extreme of having the convection scheme switched off. This appears to be an attractive compromise since it allows good representation of severe convection events which are the main phenomena of interest. The disadvantage of doing this will be that relatively light, scattered showers will tend to be missed altogether (or aliased into larger showers). In order to capture these events it would either be possible to postprocess the convection scheme output or even to use output from the 12km model.

The 1km gridlength model was run without a convection scheme and was generally found to produce much more satisfactory results than the 4km model. The tendency to produce too much rain is greatly reduced as is the delay in initiation. These features can be seen in Figure 2.2. The handling of convection in the 1km model also depends on the convection in the 4km model providing the boundary conditions.

Figure 2.4(a) shows the model representation of a case of scattered convection moving into the domain across the northern boundary. It is striking that, unlike in the equivalent radar picture (2.4 c), there are no showers in the northern third of the domain (this also appears to have a knock on effect in that the showers don't correctly die out in the southern third). As a result of the driving 4km model having no explicit convection in this case there are no cells being advected in through the boundaries and the convection takes a significant distance into the domain to spin up. Figure 2.4(b) shows the same field but for a 1km model driven by a 4km model with no convection scheme. As expected there is now significant convection in the northern part of the domain from convective cells being advected in through the boundaries. It is evident that the 4km models too widely spaced cells lead to a similarly unrealistic distribution of rainfall in the northern part of the 1km model. The southern half of the domain, however, contains a very similar shower distribution to the case where the driving model had no convection.

It may be concluded that there is always likely to be some kind of spin up problem at the inflow boundary but that this may not affect the rest of the domain. An analogous

problem with spinning up of convection from the inflow boundary is also seen in the 4km model (which being driven by the 12km model boundary conditions always has to spin up the convection from nothing).

The operational 12km model is run without any horizontal diffusion with the exception of the "targeted diffusion" scheme which effectively inserts a large amount of diffusion where there are strong vertical velocities in order to suppress gridpoint storms. Currently the 4km and 1km models are being run with a small amount of horizontal diffusion in order to reduce the amount of gridscale structure in a number of model fields including vertical velocity and precipitation. This has the detrimental effect of making the delay in the initiation of explicit convection even worse although it tends to make the overall rainfall rate more correct. The implication of these results is that diffusion is required where convection is actually taking place but not elsewhere. These conditions might be met by implementing a 3d turbulence scheme. Plans to implement such a scheme are currently underway for future use however in the current work a quite small (8 timestep del-4) constant amount of horizontal diffusion was applied.

Finally changes have been made to the microphysics parameterisation in the model. The currently available version of the UM assumes that rain falls straight out of the model without being advected by the winds. Consideration of typical rain fall velocities and wind speeds shows that this is likely to be poor approximation on horizontal scales of less than 10km. For this reason prognostic rain will shortly be included in the 1km model runs (although it was not in the runs reported here). As would be expected, in practice the horizontal movement of the rain represents a smoothing of the precipitation fields. This is likely to have a positive effect on the gridscale problems mentioned in the last paragraph. More importantly the advection of rain is expected to reduce the tendency of rain to fall into the updraft of showers which may act to unrealistically reduce their intensity. The prognostic rain included in the UM has been shown to have a positive impact on the verification of the model precipitation in cases of orographic rainfall.

(d) Initial data/Data assimilation

A key question with high resolution models is how to provide initial data. The simplest approach is not to attempt any high resolution data assimilation but simply to start each forecast from a lower resolution analysis (12km in this case). This method has been used by many workers and was the only approach used in the early stages of this project. It was also used in the current work in order to produce control runs against which to compare the runs with assimilation. The obvious disadvantage of this technique is that the 4km and 1km forecasts take a finite time to spin up high resolution structure (around two hours). Since the period of interest in these forecasts is only out to T+6 this represents a serious reduction in the usefulness of the forecasts. For this reason a continuous system with high resolution assimilation has been set up to enable the high resolution structure from one forecast cycle to be propagated forward into the next.

The Met Office current operational 12km model uses a 3D variational assimilation system (3D-Var) (Lorenc et al 2000) which generates a set of increments which are nudged into the model using an Incremental Analysis Update (IAU) scheme for an hour on each side of the nominal analysis time. These increments are constant in time and independent of the evolution of the model. The increments at each timestep being

the total analysis increments divided by the number of timesteps in the 2 hour period. The 3D-Var includes surface temperature, humidity, pressure, winds and visibility, radiosonde, pilot and dropsonde data, satellite atmospheric motion winds, ATOVS radiances and wind profiler data.

In addition the Moisture Observations Processing System (MOPS) provides fields of surface precipitation rates at 5km resolution and 3D cloud cover at 15km resolution derived from surface observations, radar and IR geostationary satellite imagery data via the NIMROD system. The precipitation data is reduced to 15km resolution and assimilated into the model via a Latent Heat Nudging (LHN) scheme (Jones and Macpherson 1997). The MOPS cloud data is converted to profiles of relative humidity and nudged into the model along with the latent heating increments using an Analysis Correction (AC) scheme. This nudging is carried out over a period either side of the analysis time which is T-2 to T+0.5 hours for the cloud and T-2 to T+2 for the precipitation (in the case of the latter this period overlaps with the period of nudging for the next cycle). The weights and values of these MOPS relative humidity and temperature increments vary in time and depend on the evolution of the model. The analysis is performed every 3 hours and uses 3 hourly MOPS cloud data and hourly MOPS precipitation data.

In order to run the 4km model with assimilation the 3D-Var and MOPS/LHN systems have initially been taken over from the 12km model with few changes. There are improvements that could be made by tuning some of the parameters of the existing system (for example the error covariances or the correlation lengths) as well as the incorporation of new sources of higher resolution data. Work on these aspects is in hand and will be implemented in future versions. The aim of the work described here was to evaluate the minimum benefits which might be obtained from continuous assimilation rather than an attempt to optimise the assimilation itself. It is important to realise that this technique does not add any new information on scales less than 12km so the high resolution information will still be self generated by the model. It is hoped, however, that in future high resolution data (eg radar data) will be assimilated into the model.

A second technique for the 4km assimilation has been tested which is to interpolate the increments from the 12km 3D-Var analysis to the 4km grid and use them in the 4km model via the IAU. This saves a great deal of processing time by avoiding running the full 3d-Var analysis at 4km. The danger with this technique is that if the 4km and 12km backgrounds are significantly different the 12km increments are no longer correct for the 4km model. This method will be described in more detail in a future paper and will not be discussed any further here save to remark that the results from the limited number of cases available so far appear to be comparable with those from running 3D-Var at 4km.

In the case of the 1km model it is not currently practical to run 3D-Var because of the size of the domain. The domain is large in terms of model gridpoints and it would therefore be costly to run 3D-var. More fundamentally the physical size of the domain is only a few times larger than the correlation length. This means that problems related to the boundaries (for example the fact that 3d-var assumes zero increments at the boundaries and the neglect of observations just outside the domain) become more significant. For these reasons it was considered that the only option to establish a continuous assimilation cycle was to use the increments from the 4km or 12km models and in the work shown here the 4km increments were used.

Due to the smallness of the 1km domain (in physical terms) there is evidence that the model is more influenced by the boundary conditions than by the increments. In one of the cases discussed here the suite was rerun continuously for a number of cycles without any increments being put into the 1km model and the results were only slightly different from the full run. At the time when this work was carried out it was not possible to run the MOPS/LHN system at 1km for technical reasons so the 1km results shown do not include MOPS/LHN. (More recently work has been carried out to make this possible so MOPS/LHN will be included in future 1km runs.) A second possibility, rather than running a continuous 1km model, would be to spin up each 1km forecast from the 4km analysis - it might be expected that the spin up problem would be less going only from 4km to 1km than from the 12km to the 1km especially if the convection was being handled explicitly in both the 4km and 1km models. Details of the suite including 4km and 1km assimilation may be seen in appendix B.

## 3. Cases

In order to test the forecast system in a realistic mode and gather reliable statistics a long period of data would be required. An example would be to run the models every 3 hours for several months. Unfortunately this is not possible with the current computing resources. The longest period which can comfortably be run given that a number of different configurations were being tested was around 24 hours (i.e. 8 assimilation/forecast sets) including any spin up of the assimilation before the period of interest. This greatly reduced timeframe has necessitated a return to the methodology of running particular cases chosen for their meteorological relevance.

For the study reported here four cases from the summer of 2003 were chosen. These are summarised in table 3.1. Since the empasis of this work is on convection all four cases contained elements of convection although case 4 also contained rain on somewhat larger scales. Cases 1 and 2 were chosen to represent examples of strongly and more weakly forced convection respectively. The table lists the times for which 6 hour forecasts were run with the three models – in addition 3 cycles were added before the first forecast to enable the assimilation to spin up the high resolution models. Case 4 was run for 48 hours (from QM00 on 28$^{th}$ Aug to QM21 on 29$^{th}$) simply to demonstrate that this was possible. Due to the archive system problems discussed in appendix B this case took an unacceptably long time (around a month) to run. Furthermore for consistency with other cases the statistical analysis discussed here was carried out only for the cycles QM06-QM15 on the 28$^{th}$ so only these cycles will be considered in what follows.

## 4. Overview of results

### (a) General

This section attempts to give a general overview of the results from the high resolution models for the four summer 2003 cases which are the subject of this paper. Each trial consisted of four forecasts so this overview will, of necessity, only give a few samples of the output obtained. These have been chosen to highlight the time of most interest in the particular case. A statistical analysis taking into account all the data available is given in section 5. This section gives results from the runs using

assimilation in the high resolution models, a comparison with runs spinning up from 12km analyses is given in section (b) below.

*13th May 2003*

The 13th May case consisted of scattered showers which developed starting around 08 UTC into a fairly uniform scattering of showers over the 1km domain from 10 UTC onwards. From around 14:30 onwards a line of heavy thunderstorms developed rapidly peaking at around 16 UTC. Figure 4.1 compares the instantaneous rain rates from the model forecasts started from 09 UTC data to the radar at 15 UTC. The main line of thunderstorms can be seen in the radar data (frame d) as the curved feature running between the Bristol Channel and Sussex. The 12km model produces too low intensities of rain and it is not even obvious from the precipitation field that there is a line of showers. Analysis of the total cloud fields for this model shows that there *is* a band of cloud corresponding roughly to the position of the line of showers in the 4km model. The 4km model is an improvement on this in the sense that it does give a line of heavy showers. These showers are in the wrong place but would at least alert a forecaster to the possibility of heavy rain. The 1km model looks better than the 4km to the eye in that it produces a curved line of showers roughly in the right place (slightly too far east and not extending far enough west). This line has somewhat too high peak intensities when compared to the radar but has less lower intensity around the main band. It is also evident that rain in other parts of the domain (eg to the north) is not represented in the 1km model.

For many applications (eg flood forecasting) it is most relevant to consider rain accumulations. Figure 4.2 shows a 6 hour accumulation for the same case. The 12km model fails to produce the high rain totals concentrated in relatively small areas (although when averaged over the whole domain it produces about the right amount). The 4km model does produce some swathes of heavy rain although it misses the main swathe seen in the radar data due to the storms. In contrast the 1km model does appear to have produced the storms but it is now apparent that it has, overall, produced too little precipitation from them. It is also noticeable that both the 4km and 1km model produce too much rain over the Channel. It is very noticeable that the 4km and 1km models share with the radar similar patterns of NW-SE oriented streaks which are absent in the 12km model. This is because the 4km and 1km models produce the rain as explicit showers which are advected in the NE flow as in reality. The 12km model produces rain only in the convection scheme which, by its single column nature, does not allow for advection of patches of rain.

We have also looked at the 15 UTC mean sea level pressure fields from the models shown in figure 4.3. The 4km and 1km models have much more short range structure and even when they are averaged up to the 12km grid there are features which do not appear in the 12km model. The implication of this is that the high resolution models are adding information even on the larger scales. Of particular relevance in the 1km model is the ridge of high pressure corresponding to the band of rain with a low pressure region to the south. Comparison of the fields in figure 4.3(e) to 4.4(b) (1 hour later) shows that an hour later the high/low dipole structure has advected with the flow. It is important to check how these pressure fields compare to observations. Figure 4.4 shows a hand analysis from surface pressure observations at 16 UTC compared to the equivalent model field. There is encouraging agreement in that both model and observations have a low pressure area north of the West Sussex coast with

the ridge of high pressure to the north. Despite the fact that this ridge is significantly more pronounced in the model, possibly as a result of the high local rain intensities generated in the model, the observations show that the extra structure generated in the 1km model is reasonable.

*25th May 2003*

This case included scattered showers which were distributed relatively evenly but with some organisation into lines near the south coast around 12 UTC. Figure 4.5 shows the radar field at this time compared to the model precipitation fields from the 06 UTC forecasts. The 12km forecasts produce relatively large areas of light rain whereas the 4km model produces showers which are too large, too heavy and too far apart. This is the usual behaviour when the 4km model attempts to reproduce showers on relatively small scales with explicit convection. The 1km model output looks more realistic and although wrong in detail hints at the organisation in the Sussex/Kent area. All three models give about the right amount of rainfall when averaged over the 1km domain area. Figure 4.6 shows the 07-13 UTC accumulation for the same case. This emphasises that the 1km, and to a lesser degree the 4km, model does better than the 12km in producing the rain in the Sussex/Kent area but also that all models produce too much rain towards the north west corner of the domain.

*1st July 2003*

The 1st July case had several roughly NE-SW oriented lines of convection moving south over southern England through the middle of the day. Figure 4.7 shows the precipitation rates at 12:30 UTC from the 06 UTC forecasts. The 1km model produces a linear feature but too far back. The 4km and 12km models produce features in roughly the same place but with peak rainrates too high in the 4km model. In the 12km model the rain is more spread out with much lower peak values. Figure 4.8 shows the corresponding 7-13 UTC accumulations which show that there was too much rain over too wide an area in the 4km and 1km models. Examination of the fields shows that the models produced too much rain (domain averaged) at the start of the run and approximately the right amount after about 10 UTC. This is an effect related to assimilation which is discussed in section 4(b). The excess of rain is nearly a factor of two in the 4km and 1km models.

*28th August 2003*

This case consists of larger bands of rain, oriented approximately NE-SW appearing and move northwards over south east England in a southeasterly flow. Figure 4.9 shows precipitation rates for 16 UTC from the 09UTC forecast for this case. The 12km model gives a good indication of the area of the two main rain bands although they are too far north and the intensity is too low. All this rain in the 12km model is generated by the convection scheme. The 4km model gives several intense cells produced explicitly. In addition the 4km model has light rainfall (less than 0.1 mm hr) which is not visible in the figure mapping out similar bands to those visible in the 12km model. The 4km model cloud field (fig 4.10) shows that the bands are also visible in cloud. Presumably there is not sufficient forcing for the rain to be produced by explicit convection. In contrast the 1km model does produce lines of showers along the rain bands. All the models produce around the same rainfall rates when

averaged over the 1km domain and these are around half that seen in the radar. Figure 4.11 shows the corresponding 10-16UTC accumulations. As in the 13th May case the 4km and 1km models share with the radar accumulation fields streaky patterns (in this case oriented SW-NE) resulting from the advection of the showers.

Although these four cases are all different some general comments can be made from examining the results:

- The 12km model tends to badly underestimate the peak intensity during convective events however in most cases it produces enough or too much rain when averaged over the whole 1km domain area. The 4km and 1km models produce much higher peak intensities which generally represent a great improvement but may be too high in some cases. In cases where the peak intensities are too high in the 1km model this does not necessarily lead to too high accumulations because there is often a corresponding lack of rain away from the peak.
- In cases of larger showers the explicit convection in the 4km model often gives an improvement over the 12km since the showers are able to organise and advect in a more realistic manner than with the convection scheme. This can result in the overall area of precipitation being better than in the 12km model. The less positive side of representing most of the convection explicitly is that in situations where the showers are too small to be resolved, the 4km model has the tendency to produce rain in too large, too intense and too widely separated showers.
- The 1km model often produces the most realistic showers in terms of area covered by showers, size and peak intensity of showers. The details of the individual showers are often wrong but this is to be expected from predictability arguments. The 1km model shares the advantages that the 4km model obtains from attempting to represent the convection explicitly but suffers from less of the disadvantages resulting from the grid length being too large to represent smaller showers.
- If the large scale structure is incorrect in the 12km model it is unlikely that running high resolution models will improve the situation.

It is worth mentioning that we have observed other cases (e.g. frontal rain) where the 4km and 1km produce virtually identical results but these results are still an improvement over the 12km model.

(b) Effect of assimilation

In this section results are presented comparing the results from the high resolution models including assimilation to those spinning up each forecast from the 12km model. A statistical comparison is presented in section 5.

As discussed in section 2(c) the major disadvantage of spinning up each high resolution forecast from a 12km analysis is that it will be unusable for precipitation prediction for the first few hours due to the need to spin up realistic structures. In convective situations the 12km model will be representing the rain in the convection scheme so the time is required to start the explicit convection in the high resolution model. An example of this is shown in figure 4.12 where the T+2 forecast from the

model with continuous assimilation gives a good indication of the thunderstorms whereas the forecast spinning up from the 12km T+1 analysis has virtually no rain.

Although not as immediately obvious we also find that the rainfall starting out wrong in the spin up forecast has a negative impact even after 6 hours. Figure 4.13 shows the 1km area domain averaged hourly accumulations as a function of forecast time averaged over all 16 forecasts used in this study. As expected the 4km and 1km models spinning up from the 12km analysis produce too little rain at the start of the forecast. However at later times they produce too much rain, apparently peaking after 5 hours. It is impossible from the current work to say what happens after longer times due to only 6 hour forecasts having been run. Figure 4.14 shows the domain averaged rain rates for one forecast and the corresponding domain averaged CAPE values. The too low rainfall rate at the start of the spin-up forecast appears to lead to a more rapid build up of CAPE which is then consistent with the overshoot in rain rate later on. In contrast the equivalent run with assimilation has approximately the correct rain rate at the start which releases the CAPE and avoids the build up which leads to too much rain later on.

Turning back to the rates averaged over all the runs, figure 4.13 shows that runs with continuous assimilation cycles (at all three resolutions) tend to produce roughly the correct amount of rainfall for the second three hours of the forecasts but too much rain at the start of the forecast. It has been noticed in many individual forecasts that in a convectively unstable situation there is a tendency for there to be much more rain at the end of an assimilation cycle than in the forecast for the equivalent time from the previous cycle. It appears that one effect of the assimilation is to encourage initiation of convection (often spurious). This will be a subject for future work however it is possible that the assimilation simply adds noise which encourages initiation. The MOPS/LHN system which is intended, in part, to add precipitation information to the model is also likely to be relevant to this question.

The conclusion from this comparison is that some form of assimilation system will be essential for a short range high resolution forecast models in order to avoid the problem of too little rain early in the forecasts and too much later on. Future work will determine whether this will be best achieved by a truly continuous system or simply by starting each run further back than required and running for a period with assimilation before the forecast starts.

## 5. Statistical analysis

In this section verification statistics for the cases run are presented. As well as providing a more rigorous assessment of the forecasts this has the advantage that it is possible to compound results from all the forecasts rather than presenting snapshots as in section 4. In section (a) we discuss results from RMS and bias statistics compared to observations for various surface quantities and in (b) precipitation verification statistics are presented.

A general problem when assessing high resolution models is how to take account of the frequently observed situation of features being generated correctly but in the wrong place. This may happen, for example in the case of a front, if the larger scale driving model is wrong. Alternatively in a situation with scattered showers it may happen simply because the predictability time is short for small features. It is very unlikely that in a field of small, scattered showers (such as in the 25$^{th}$ May case) that the location of every cell will be correct. If, however, the area covered by the showers

and their intensity and distribution is correct then the forecast will still be very useful. For these reasons it is desirable to use verification methods which can, at least optionally, take account of small displacement errors of features.

(a) Surface Statistics

These were obtained using the Met Office Area Based Verification System which was used to generate statistics comparing surface observations to the model. Quantities compared were surface pressure, temperature, winds, humidity, visibility and cloud cover. The RMS error and biases were calculated for each of these although only the first two are presented here. The values of the observations were compared to model values calculated by interpolation of the four gridpoints closest to the observation position. The data shown here were calculated by taking data on the original model grids to calculate the statistics. The statistics were also calculated after aggregating data from the 4km and 1km models onto the 12km grid in order to make the comparison as fair as possible. In practice this was found to make only very small differences to the results. The temperature observations were not corrected for height. These statistics are of limited usefulness for assessing the high resolution models partly because they do not include precipitation which is the key field in this work. A more fundamental problem is that RMS statistics tend to penalise high resolution structure which is in the wrong place. A model which generates a feature correctly but in the wrong place will get a worse RMS score than one that doesn't generate the feature at all.

Figure 5.1 shows the RMS error and biases for the mean sea level pressure and temperature averaged over the four cases comparing the 3 model resolutions including 4km and 1km forecasts run both with assimilation and spinning up each cycle from 12km analyses. The bias and RMS error in the pressure and temperature data for the 4km and 1km models running with assimilation are comparable to those seen in the 12km model (taking into account the approximate spread indicated by the error bars). The implication is that the high resolution models don't give significantly worse results by this measure than the 12km model. This is the most important result from this verification method.

It is particularly encouraging that this is the case for the RMS statistics given the likelihood that increased high resolution structure would tend to lead to larger RMS errors. It is also evident in figure 5.1 that the RMS errors and bias tend to get worse as the forecast length increases which is what would be expected. The 4km and 1km models spinning up from 12km analyses give similar but somewhat worse RMS errors except near the start of the forecasts. It is to be expected that the RMS errors from the spin up forecasts near the start will be lower due to them having smoother fields. The biases of these models tend to be somewhat different to and mostly worse than those from the models with assimilation however given the relatively small sample size this has not been analysed in detail.

(b) Precipitation Statistics

The precipitation statistics evaluate the models precipitation fields against data from the Met Office radar network. These have been designed with the idea of taking into account varying degrees of small range spatial error as discussed at the start of the section. They are described in more detail by Roberts (2004). The standard method of producing categorical statistics is to carry out a point by point comparison of corresponding points in the radar and model data using binary 1 if the threshold is exceeded and 0 otherwise. In order to take account of spatial error the binary model

data is replaced with a probability obtained by counting the proportion of points above the threshold within a sampling radius of each point. This represents ignoring all spatial information within the sampling radius. In some cases the same procedure is also used for the radar data. The resulting statistics can then be plotted as a function of sampling radius in order to give an impression of the effect of neglecting spatial errors of various sizes.

Two statistics are discussed here. The first is a fraction skill score which gives information about the intensity and the distribution of the precipitation and is calculated comparing the model and radar data both converted into probabilities. This ranges from 1 for a perfect forecast to values below 0 for a very poor one. The second is a Brier skill score which gives information about the precipitation distribution and is calculated from the model data converted into probabilities compared to the point by point data from the radar. The range of this is from 1 for a perfect forecast to 0 for a poor one. More detailed descriptions and definitions of these two statistics may be found in Appendix A. This analysis is carried out on a 5km grid over the 1km domain area (the 1km and 4km data are area averaged onto this grid and the 12km data interpolated). The 5km gridlength is chosen because that is the highest resolution of the available radar data over the 1km domain area.

Figure 5.2 shows the fraction skill scores for each of the four cases aggregated over the four forecasts as a function of sampling radius. These results are for a threshold of 4.0mm accumulated over the 6 hours of the forecast from T+1. As would be expected the values of this score increase as the sampling radius increases and larger spatial errors are ignored. For two of the cases the 12km model has very low scores – this is because it largely fails to produce any rainfall accumulations over the threshold. It is obvious that the scores from the various models are very different between the four cases. This means that in order to get robust statistics by aggregating data from many cases it would be necessary to include many more than four cases. Figure 5.3 which shows the same curves including the aggregated results from the four cases aggregated together should therefore be treated with extreme caution. Despite this caveat it is very encouraging that the high resolution models appear to outperform the 12km models at all sampling radii. It is also evident that the runs including assimilation do approximately as well as the spin up runs (except at the shortest sampling radii). Given that the assimilation is essential for short range forecasts but has not yet been optimised this is also encouraging.

These fraction skill scores have also been calculated for a much lower accumulation threshold, 0.001mm (not illustrated) and also for a higher one, 8.0mm (figure 5.4). In the 0.001mm case the curves from all the models virtually lie on top of each other and have much higher values (0.8 at zero sampling radius increasing to around 0.95 for longer values). This is unsurprising in that the very low rain amounts are well predicted by all the models and that the high resolution models add very little. The 8.0mm curves give the 12km model a very low score because the 12km model only rarely produces rainfall of this magnitude. The same limitations about the small number of cases apply as in the 4.0mm case but more extreme because there are less points with these higher accumulations. It is again encouraging to see the models with assimilation (especially the 1km) doing well compared to the ones spun up from 12km analyses.

The 6 hour accumulations have also been analysed in terms of the Brier skill score as shown in figure 5.5 and 5.6. These scores have been calculated using a precipitation threshold which varies with forecast and model rather than being fixed. This threshold

is chosen so that it includes the 10% of pixels in the grid with the highest rainfall. The score therefore gives information about the spatial structure of the rainfall rather than the overall intensity. As shown in appendix A this score will always tend to a value of 0.5 for the largest sampling radii – the interest is in where the curve first crosses the 0.5 line and where it reaches a maximum which give information on the spatial accuracy of the forecasts. The former measures where the forecast is better than random and the latter the value at which the forecast is most skilful. As with the fraction skill scores figure 5.5 shows that there is a great deal of variability between cases and so the overall aggregated Brier skill scores in figure 5.6 should be treated with caution. However there does seem to be a consistent signal that the spatial error tends to be smaller in the 4km and 1km forecasts (the peak of the curves is at around 50km) than in the 12km ones (70km). It also appears that the spatial error is smaller in the forecasts spun up from 12km analyses (30km) than those with high resolution assimilation.

The 6 hour accumulations cover the whole length of the forecasts. It is also desirable to get some information as a function of forecast length particularly in the light of the expected benefits of assimilation for short forecasts ranges. Figures 5.7 and 5.8 show the fraction skill scores calculated from hourly accumulations with the threshold set to 1mm plotted as a function of forecast range. In order to produce this plot it is necessary to choose a fixed value of the sampling radius. A radius of 55km has been chosen as representative since at this value most of the forecasts have a reasonable amount of skill. It is noticeable that the overall values of these skill scores are lower than those for 6 hour accumulations. This is likely to be because taking into account only one hour leads to timing errors being more detrimental to the forecast. The variability of the statistics from case to case is once again very obvious and it is also noticeable that many of the curves show large variations from one hour to the next. In the light of this it is difficult to draw firm conclusions from the curves aggregated from all the cases (figure 5.8). The 12km model is the only one which shows a strong signal of the skill reducing with forecast length. This may be related to this being the only model running with the standard UM convection scheme which, as can be seen in figure 2.2 tends to die out convection too quickly. As a result towards the end of the forecasts the 1km and 4km models offer the biggest benefit over the 12km model. The 4km and 1km runs spinning up from 12km analyses appear to be slightly worse at the start.

To summarise; the precipitation verification statistics which have been calculated confirm the subjective conclusion that the 4km and 1km models provide benefit over the 12km. Additionally the models with assimilation at 4km and 1km seem to perform approximately as well for intermediate rainfall thresholds and better for higher thresholds compared to those spinning up despite the fact that the assimilation has not yet been optimised. The exception to this is for shorter sampling radii (below 60km) where the spin up runs appear to be better. The caveat to these conclusions is that only a relatively small number of cases have been run (and each case gave significantly different results) so that these statistics are not very robust.

## 6. Conclusions and Future Work

### (a) Current status

The HRTM, a prototype high resolution forecast system based on the Met Office non-hydrostatic UM, has been constructed and tested on cases over summer 2003. The

primary motivation for this model has been forecasting of extreme convective events out to 6 hours. This system includes 4km and 1km models as well as a 12km model for comparison (and to provide the boundary conditions). The 1km model does not use a convection scheme and the 4km model has a modified scheme to encourage the model to represent severe convection explicitly. In order to make feasible useful short range forecasts (up to 2 hours) continuous assimilation cycles have been introduced into the 4km and 1km models. This is currently based on the philosophy of assimilating low resolution information and allowing the high resolution structures to propagate from one forecast to the next.

In this paper we have reported the tests of this system on four convective rainfall cases from summer 2003 with each case consisting of four forecast runs. Objective verification against surface station data has been carried out for mean sea level temperature and temperature with the conclusion that the biases and RMS errors are not greatly different in the high resolution models from the 12km one.

The precipitation forecasts have been analysed both by subjectively assessing the model fields and with objective scores. Subjective analysis shows that the 4km and 1km models can provide greatly improved forecasts for higher rainfall events. The objective scores provide encouraging evidence that the 4km and 1km models do better at forecasting the higher precipitation events and that the continuous assimilation successfully enables useful forecasts to be produced at short forecast ranges. For intermediate rainfall amounts (4mm over 6 hours) the statistics do not provide evidence that the 1km model produces benefits over the 4km however for higher thresholds (8mm over 6 hours) the 1km model does appear better. These scores were calculated for a number of sampling radii. It was found that whereas the 12km model is most accurate when spatial errors of around 70km are neglected, for the 4km and 1km models with assimilation the equivalent scale is 50km. The 4km and 1km models spinning up from 12km analyses seem to be accurate over 30km scales. The reason why these runs are better in this sense than the assimilation ones will be the subject of future investigation.

(b) Next Phase of HRTM Project

The encouraging results presented here need to be checked in a more representative fashion by carrying out longer trials. There are also numerous other improvements to the system under consideration. These are summarised in this section which presents the main aims for the next phase of the project.

The next phase of the project (until March 2005) will be centred around a longer trial based on cases from summer 2004. There have been more convective cases over summer 2004 compared to 2003 so this should make the choice of cases easier. The aim will be to run more cases than the four which were possible in the current work. This will be made possible by the use of the new NEC supercomputer and also by improvements to the postprocessing system which will make the gathering and aggregating of statistics easier. In addition to gathering statistics attempts will be made to gain more understanding of model behaviour by analysing or more cases in more detail.

Specific model problems which will be addressed will include:

- Improvements to the convective and turbulence parameterisations. A shallow convection scheme will be tested.

- Improvements to the 3d-Var assimilation system at 4km – currently the system has not been optimised for high resolution use and is using settings for the 12km model. The current system suffers, due to the small domain size from the fact that Var puts the increments to zero at the edge of the domain. Future work will investigate scale selective Var analysis where the analysis is only carried out for shorter scales and larger scale increments are taken from the 12km analysis.
- Implementation of the MOPs precipitation and cloud assimilation in the 1km model. This will introduce the first source of high resolution data into the models since precipitation data on a 5km grid will be used for input.

(c) Longer Term

In the longer term there are a number of areas for improvement to the models including:

- High resolution assimilation (for example of radar data).
- An ensemble system in order to capture the uncertainty from the lack of predictability on short scales.
- Variable resolution models rather than simply nesting one model inside another.

References.

Cullen, M.J.P., Davies, T., Mawson, M.H., James, J.A., Coulter, S.C. and Malcolm A., 1997, An overview of Numerical Methods for the Next Generation UK NWP and Climate Model" Numerical Methods in Atmospheric and Ocean Modelling. The Andre J.Robert memorial volume. Edited by Charles A Lin, Rene Laprise and Harold Ritchie 425-444

Gregory, D. and Rowntree, P.R. 1990, A Mass Flux Convection Scheme with Representation of Cloud Ensemble Characteristics and Stability-Dependent Closure. Mon. Wea. Rev. **118** 1483-1506

Jones, C.D. and Macpherson, B. 1997 A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. Meteorol. Appl. **4**, 269-277

Lean, H.W. and Clark, P.A. 2003 The Effects of Changing Resolution on the Mesoscale Modelling of Line Convection and Slantwise Circulations in FASTEX IOP16 Quart. J. Roy. Meteor Soc.,**129** 2255-2278.

Lock, A.P., Brown, A.R., Bush, M.R., Martin, G.M. and Smith, R.N.B. 2000, A New Boundary Layer Mixing Scheme. Part 1: Scheme Description and Single-Column Model Tests. Mon. Wea. Rev. **128** 3187-3199

Lorenc, A. et. al 2000., The Met Office Global 3 Dimensional Data Assimilation System Quart. J. Roy. Meteor Soc., **126**, 2991-3012

Roberts, N.M., 2003, The impact of a change to the use of the convection scheme to high resolution simulations of convective events. FR tech report 407/JCMM report 142.

Roberts, N.M., 2004 Measuring the fit of rainfall analyses and forecasts to radar: Strategy for stage 5 evaluation Stage 4 report from the Storm-Scale numerical modelling project. FR tech report 432/JCMM report 146.

Wilson, D. R., Ballard, S. P., 1999. A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. Quart. J. Roy. Meteor. Soc., **125**, 1607-1636.

Appendix A

Description of Precipitation Statistics.

This appendix gives more details of the precipitation statistics used in this report. The interested reader is referred to Roberts (2004) for more information. In order to carry out the calculation of the statistics the model rainfall data is first interpolated or averaged onto the 5km grid with the same area as the 1km domain. It is then converted into a probability by searching a square around each gridpoint of side length of the sampling radius and evaluating the number of gridboxes with a rainfall accumulation above the threshold. In order to evaluate the forecasts this probability data is compared with the radar data and there is a choice about how to do this. One option is to take the radar data at each point simply as a binary 1 or 0 depending on whether the threshold was exceeded or not. A second option would be to convert the radar data into probabilities in the same way as the model data. The first approach seems the most natural rather than expressing the radar data as probabilities and it is also the usual definition of a Brier score. The problem is that using the model data as probabilities and the radar as binary introduces an undesirable asymmetry in the sense that interchanging the radar and model fields would give a different score. This is not a problem if the model and radar data are renormalized by using relative rather than absolute rainfall thresholds. For example a threshold might be chosen separately for the radar and rainfall data such that the points in the domain above the threshold represent 10% of the total points in each case. For this reason the binary radar data is used to calculate a Brier skill score with these relative thresholds which is primarily used to give information about the spatial distribution of the rainfall. It is also necessary to have a score which takes account of the absolute amount of rainfall. In order to get around the asymmetry problem this is calculated with the radar converted into probabilities in the same way as the model. This is referred to as the "fraction skill score" in this paper. The definitions of these two scores are given below.

1. Brier Skill Score

This is calculated by comparing the model probability of exceeding the chosen threshold with the radar data. In this case the radar data is not converted to a fraction but is either 1 or 0 depending on whether the threshold was exceeded at that gridpoint. As discussed above the precipitation thresholds used here are relative (eg top 10% of points) rather than absolute. The Brier score is defined by

$$BS = \frac{1}{N}\sum_{i=1}^{N}\left(O_i - f_i\right)^2 \qquad \text{(A1)}$$

where $f_i$ is the fraction from the forecast and $O_i$ is now the binary data from the radar. The perfect score is still zero since ideally the forecast would give zero probability where rain was not observed above the threshold and one where it was. The problem with this is that if there is only a small area in the domain with rain accumulation over the threshold then the score will be good by virtue of the large number of points where both the forecast probability and radar data give zero. In addition in this situation the score will be better for a model which doesn't produce the rain at all than in one that produces the rain but in the wrong place. In order to avoid these problems a second score called the Brier skill score is used defined by:

$$BSS = 1 - \left( \frac{BS}{2 f_0 (1 - f_0)} \right) \qquad \text{(A2)}$$

where BS is the Brier scores defined as in (A1) and $f_o$ is the relative frequency threshold. The denominator of the fraction in brackets here is the Brier score for a randomly distributed field with frequency $f_o$. The best value of the Brier skill score is 1.0 and forecasts with worse than random skill will give values below zero.

In the current paper the Brier skill score is plotted against sampling radius. The precipitation threshold is set so that the number of gridpoints exceeding it is fixed at 10%. As the sampling radius is increased from zero it is expected that the score will initially increase due to spatial errors in the forecast being ignored. However beyond a certain point the score will fall again due to the useful spatial information being lost. For large enough sampling radius all spatial information is lost so that the score will tend to 0.5 (due to the factor of 2 in the denominator of (A3). There will be a peak in the curve above 0.5 for good forecasts however for poor forecasts the score will stay below 0.5 for all values of the sampling radius and there will be no peak. Where there is a peak the sampling radius at which this peak occurs give information on the spatial accuracy of the forecasts. The value of 0.5 represents the point at which the forecast is better than simply predicting a probability of $f_o$ everywhere. Hence an alternative measure of the spatial accuracy of the forecasts is the sampling radius at which the curve first exceeds 0.5.

2. Fraction Skill Score

In order to calculate the skill score the radar data is also converted into probabilities in exactly the same way as the model data using a rainfall threshold which may be absolute rather than relative. The definition of the fraction skill score is given by:

$$FSS = 1 - \left( \frac{\sum_{i=1}^{N} (O_i - f_i)^2}{\left( \sum_{i=1}^{N} O_i^2 + \sum_{i=1}^{N} f_i^2 \right)} \right) \qquad \text{(A3)}$$

where $O_i$ is the observed fraction from the radar data and $f_i$ is the forecast fraction. This score is essentially the root mean squared error in the probabilities normalised by the error which would be seen if the forecast and observed errors were completely uncorrelated (i.e. if wherever the forecast probablility was non-zero the observed probability was zero and vice versa). The best possible score would be if the numerator of the expression in brackets was zero so the skill score would be 1.0. The worst possible score would be if the numerator of the expression in brackets is equal to the denominator which gives a score of zero.

Appendix B

Details of Suites

In this appendix more details are given of the suites used to generate the results presented in this paper. As described elsewhere each case was run with a suite including data assimilation in the 4km and 1km models and also with one simply spinning up the 4km and 1km models from the 12km T+1 analysis each cycle. The 12km model was also run in addition to the 4km and 1km ones both to enable comparisons with the high resolution models and in order to provide boundary conditions for the 4km run.

As described in section 2(d) the suite including high resolution assimilation included 3d VAR and MOPs at 4km and reconfigured 4km increments at 1km. The detailed functionality of this suite assimilation is given in table B1. The 12km model was run including rerunning 3dVAR and MOPS. Although it is possible to run the 12km model without rerunning VAR or MOPS by using data from the rerun archive this was not done here partly to enable experiments with MOPS switched off to easily be performed in a consistent manner. For technical reasons (the lack of background files from the previous run) it proved to be necessary to run the first cycle without assimilation (both VAR and MOPS) in the 4km model (the same was done in the 12km model for consistency and in the 1km model since no increments were available in the first cycle). This will be potentially detrimental to the results and will negate any attempts to compare results in detail to operational runs. It would, however, be hoped that for runs of a number of cycles the missing of some data at the very start will have only a small impact.

In practice a major problem was fitting this suite into the queues available on the t3e supercomputers. Table B2 shows timing data for the models. The high resolution models require much more CPU as a result of both larger domains and the need to run with shorter timesteps. The majority of the CPU time used by the suite is required for the 1km runs which require a factor of 10 more CPU than the 4km model. The final row of table B2 is included to show that if the timestep is halved in the 1km model the increase in CPU requirement is partly offset by the reduction in the solver iteration count. Initially the suite was run as shown in table B1 with the 1km run split into two. The main suite ran the 12km and 4km models including assimilation and these models ran for 9 hours each (3 hours assimilation cycle plus 6 hour forecast). In addition the main suite ran the 1km assimilation steps (the reconfiguring of the 4km increments) and the 3 hour assimilation cycle. The 1km 6 hour forecast (along with the verification) was then spawned as a separate job. Due to the dominance of the 1km model in terms of CPU requirements, this spawned job actually required more resources than the main suite.

Jobs set up in this way ran exceeding the queue CPU limits but there were major problems with reliability and this was compounded by the fact that the SCS does not provide a way to restart spawned jobs once they fail. Jobs tended to fail due to MetDB extract failures, MASS problems affecting the archiving and other random system problems. Another problem was that running the suite in this way led to some repetition since sometimes it was required to rerun making changes to only the 4km or the 1km models which necessitated rerunning the outer models again. For this reason most of the results shown in this paper were produced using a split suite in which the 12km and 4km models were run as one suite archiving all the model files required to run the 4km and 1km models. The 1km model was then run as a separate suite using

these archived files. This suite then did not require any spawned jobs. A separate suite was then written to carry out the 4km and 1km runs spinning up from the 12km analyses again using the archived files from the 12km/4km suite. This approach also circumvented a problem which was sometimes seen with the 12km run not being reproducible. This was probably because the MetDB did not always return identical observations for an unknown reason although it is also the case that Var is not written to be reproducible. Unfortunately serious problems were encountered with this approach because the archiving system, MASS, went through a period of extreme unreliability while these runs were being carried out which made progress very slow.

In future on the new NEC supercomputer it is hoped that these problems will be ameliorated by the extra power of the machine (although this obviously means we will be trying to run larger models!). The ideal would be to have large enough queue limits so that the whole 12km, 4km and 1km suite could be run in one queue without the need to spawn the 1km separately. Initial work on the NEC machine implies that this will be possible. Even if MASS becomes more reliable it would still be desirable for testing purposes to run with the 12km and 4km split from the 1km simply to avoid duplicating runs.

| | 12km (Oper) | 4km | 1km |
|---|---|---|---|
| **Timestep** | 5 min | 60s[+] | 30s |
| **Levels** | 38 | 38** | 76 |
| **Domain** | 146x182 | 190x190 | 300x300 |
| **Horizontal Diffusion** | None | 16 tstep del-4 | 16 tstep del-4 |
| **Prognostic Rain** | NO | NO* | NO* |
| **Convection Scheme** | Std Mass flux | CAPE Dependent CAPE closure timescale | NONE |
| **Assimilation** | 3D Var and MOPS | 3D Var and MOPS or spinup from 12km | Use 4km increments or spinup from 12km or 4km |

Notes:

*Prognostic rain was not included in the runs reported here but will be in all future work

.[+] Future work will use a 100s timestep in the 4km model.

** Early work used 76 levels in the 4km model.

Table 2.1 Differences in configurations of 4km and 1km models from operational 12km model.

| Case no | Date | Model Runs (start time UTC) | Description |
|---|---|---|---|
| 1 | 13th May | 6,9,12,15 | Line of Thunderstorms |
| 2 | 25th May | 6,9,12,15 | Scattered Convection |
| 3 | 1st July | 6,9,12,15 | Line of convection |
| 4 | 28th Aug | 6,9,12,15 | Bands of convective rain |

Table 3.1 Cases from summer 2003 used in current study.

| **Task Name** | **Section** | **Run** | **Input files** | **Output Files/*directories***<br>*(directories in italics)* | **Comments.....** |
|---|---|---|---|---|---|
| Deachive prev dump. | 1 | QM00 | (from MASS oper or prev run archive) | ${IDH}_QM00_OPR_12km.T+1 | OPR in the T+1 filename indicates that this is the operational file which is at a different UM version. It is subsequently reconfigured to the current version. |
| Dearchive 12km lbcs | 1 | QM00 | (from MASS oper or prev run archive) | ${IDH}_QM00_STD_12km.alabcou1 | This task only runs on main runs in order to retrieve lbcs for next cycle. If this is an intermediate run this task is switched off since the lbcs for the next cycle will be retrieved at the start of that cycle. |
| 12km rcf | 1 | QM00 | ${IDH}_QM00_OPR_12km.T+1 | ${IDH}_QM00_STD_12km.T+1 | The reconfiguration reconfigures the T+1 dump to the current version (this is then used to start the 12km run in the first cycle). |
| 4km reconfiguration | 1 | QM00 | ${IDH}_QM00_STD_12km.T+1 | ${IDH}_QM00_STD_4km.T+1 | Reconfigure 12km T+1 dump to 4km (no reconfiguration in main cycle). |
| 1km reconfiguration | 1 | QM00 | ${IDH}_QM00_STD_4km.T+1 | ${IDH}_QM00_STD_1km.T+1 | Reconfigure from 4km rather than 12km so that orography is consistent with lbc files. |
| Create empty 12km, 4km acobs | 1 | QM00 | (copied from qrempty.ms.acobs) | *${IDH}_QM00_STD_12km.acobs*<br>*${IDH}_QM00_STD_4km.acobs* | Copy empty acobs file for use in 12km and 4km run (first cycles) |
| (Dearchive lbcs) | 2 | QM03 | (from MASS oper rerun archive) | (${IDH}_QM00_STD_12km.alabcou1) | lbc file is only dearchived on main runs (ie 00,06,12,18) - it would not be in this case hence why the file is shown in brackets. |
| Copy empty 12km acobs | 2 | QM03 | *${IDH}_QM00_STD_12km.acobs* | *${IDH}_QM03_STD_12km.acobs* | Duplicate last run acobs so that have two dummy acobs for current 12km run. |
| 12km forecast FASTRUN | 2 | QM03 | ${IDH}_QM00_STD_12km.T+1<br>${IDH}_QM00_STD_12km.alabcou1 | ${IDH}_{DATE}QM03_STD_12km.pp0-5<br>${IDH}_QM03_STD_12km.pp_003<br>${IDH}_QM03_STD_12km.T+1<br>${IDH}_QM03_STD_12km.T+3<br>${IDH}_QM03_STD_12km.umback/T+3<br>${IDH}_QM03_STD_4km.alabcou1 | FASTRUN set to true which means IAU disabled, although model otherwise behaves as if its on, so don't need anal_inc but still have AC on hence provide empty acobs files. T+3 file is dumped although not used here for consistency with reconfiguring increments suite which needs this. |
| Copy empty 4km acobs | 2 | QM03 | *${IDH}_QM00_STD_4km.acobs* | *${IDH}_QM03_STD_4km.acobs* | Duplicate last run acobs so that have two dummy acobs for current 4km run. |
| 4km forecast FASTRUN | 2 | QM03 | ${IDH}_QM00_STD_4km.T+1<br>${IDH}_QM03_STD_4km.alabcou1<br>*${IDH}_QM00_STD_4km.acobs*<br>*${IDH}_QM03_STD_4km.acobs* | ${IDH}_{DATE}QM03_STD_4km.pp0-5<br>${IDH}_QM03_STD_4km.pp_003<br>${IDH}_QM03_STD_4km.T+1<br>${IDH}_QM03_STD_4km.T+3<br>${IDH}_QM03_STD_4km.umback/T+3<br>${IDH}_QM03_STD_1km.alabcou1 | FASTRUN<br>In addition to vn4.0 produce T+3 analysis for use in reconfiguring incs to 1km and 1km lbc file. |
| 1km forecast FASTRUN (3hrs) | 2 | QM03 | ${IDH}_QM00_STD_1km.T+1<br>${IDH}_QM03_STD_1km.alabcou1 | ${IDH}_{DATE}QM03_ASS_4km.pp0-5<br>${IDH}_QM03_STD_1km.T+1 | FASTRUN. |
| **Dearchive mops and lbcs** | **2** | **QM06** | **(from MASS oper or prev run archive)** | **${IDH}_QM06_mops**<br>**${IDH}_QM06_STD_12km.alabcou1** | lbc file **is** deachived on this cycle since main run.<br>The mops data is nimrod files used to generate the acobs files. These files are independent of resolution which is why no resolution is given in the filename. Since the OPS is now being run at 12km the increments file and the acobs files are no longer dearchived. |
| **12km Create varobs and** | **2** | **QM06** | **${IDH}_QM03_STD_12km.pp_003**<br>**${IDH}_QM06_mops** | **${IDH}_QM06_STD_12km.varobs**<br>**${IDH}_QM06_STD_12km.acobs** | |

| | | | | | |
|---|---|---|---|---|---|
| **acobs** | | | | | |
| **12km Cx create** | **2** | **QM06** | **${IDH}_QM06_STD_12km.varobs**<br>**${IDH}_QM03_STD_12km.pp_003** | **${IDH}_QM03_STD_12km.var_cx** | |
| **12km Configure LS state** | **2** | **QM06** | **${IDH}_QM03_STD_12km.umback/T+3** | **${IDH}_QM06_STD_12km.LS** | |
| **12km VAR analysis** | **2** | **QM06** | **${IDH}_QM06_STD_12km.LS**<br>**${IDH}_QM06_STD_12km.varobs**<br>**${IDH}_QM03_STD_12km.var_cx** | **${IDH}_QM06_STD_12km.anal_inc** | |
| **12km forecast run** | **2** | **QM06** | **${IDH}_QM03_STD_12km.T+1**<br>**${IDH}_QM06_STD_12km.alabcou1**<br>**${IDH}_QM06_STD_12km.anal_inc**<br>**${IDH}_QM03_STD_12km.acobs**<br>**${IDH}_QM06_STD_12km.acobs** | **${IDH}_{DATE}QM06_12km.pp0-5**<br>**${IDH}_QM06_STD_12km.T+1**<br>**${IDH}_QM06_STD_4km.alabcou1** | |
| **4km Create varobs and acobs** | **2** | **QM06** | **${IDH}_QM03_STD_4km.pp_003**<br>**${IDH}_QM06_mops** | **${IDH}_QM06_STD_4km.varobs**<br>**${IDH}_QM06_STD_4km.acobs** | |
| **4km Cx create** | **2** | **QM06** | **${IDH}_QM06_STD_4km.varobs**<br>**${IDH}_QM03_STD_4km.pp_003** | **${IDH}_QM03_STD_4km.var_cx** | |
| **4km Configure LS state** | **2** | **QM06** | **${IDH}_QM03_STD_4km.umback/T+3** | **${IDH}_QM06_STD_4km.LS** | |
| **4km VAR analysis** | **2** | **QM06** | **${IDH}_QM06_STD_4km.LS**<br>**${IDH}_QM06_STD_4km.varobs**<br>**${IDH}_QM03_STD_4km.var_cx** | **${IDH}_QM06_STD_4km.anal_inc** | |
| **4km forecast** | **2** | **QM06** | **${IDH}_QM03_STD_4km.T+1**<br>**${IDH}_QM06_STD_4km.alabcou1**<br>**${IDH}_QM03_STD_4km.acobs**<br>**${IDH}_QM06_STD_4km.acobs**<br>**${IDH}_QM06_STD_4km.anal_inc** | **${IDH}_{DATE}QM06_STD_4km.pp0-5**<br>**${IDH}_QM06_STD_4km.pp_003**<br>**${IDH}_QM06_STD_4km.T+1**<br>**${IDH}_QM06_STD_4km.umback/T+3**<br>**${IDH}_QM06_STD_1km.alabcou1** | 1km boundaries also generated |
| **Create 1km incs part a Add 4km incs to analysis** | **2** | **QM06** | **${IDH}_QM03_STD_4km.T+3**<br>**${IDH}_QM06_STD_4km.anal_inc** | **${IDH}_QM06_STD_4km.anal** | **Do this for nominal analysis time for this cycle i.e. 6Z**<br>**Add the fields in the two input files using VarScr_UMFileUtils** |
| **Create 1km incs part b: Reconfigure analysis to 1km** | **2** | **QM06** | **${IDH}_QM06_STD_4km.anal** | **${IDH}_QM06_STD_4km.recon_anal** | |
| **Create 1km incs part c: Reconfigure background to 1km** | **2** | **QM06** | **${IDH}_QM03_STD_4km.T+3** | **${IDH}_QM03_STD_4km.recon_bg** | |
| **Create 1km incs part d: Subtract to create incs** | **2** | **QM06** | **{IDH}_QM06_STD_4km.recon_anal**<br>**${IDH}_QM03_STD_4km.recon_bg** | **${IDH}_QM03_STD_4km.recon_inc** | **Subtract the fields in the two input files.** |
| **1km Assimilation run (3hrs)** | **2** | **QM06** | **${IDH}_QM03_STD_1km.T+1**<br>**${IDH}_QM06_STD_1km.alabcou1**<br>**${IDH}_QM03_STD_4km.recon_inc** | **${IDH}_{DATE}QM06_ASS_1km.pp0-5**<br>**${IDH}_QM06_STD_1km.T+1** | **3 hour run time** |
| **1km forecast** | **2** | **QM06** | **${IDH}_QM06_STD_1km.T+1**<br>**${IDH}_QM06_STD_1km.alabcou1** | **${IDH}_{DATE}QM06_STD_1km.pp0-5** | **6 hour forecast. Runs only after a certain cycle is reached defined by the variable FCRUNID.** |

Table B1. Sequence of operations in suites with assimilation. This table is intended to give an indication of the sequence of operations and there is not a one to one correspondence between lines and suite tasks. Only UM/VAR operations are included (the archiving of results, tidying up files and VER tasks are omitted). The section column refers to the SCS section, section 1 runs before the cycling starts and section 2 is cycled. For concreteness it gives and example of a run which is requested to start at the QM03 cycle (i.e. section 1 runs for QM00). The section in bold type repeats until the last cycle requested is reached.

| Model | Domain | Timestep | Typical Solver Iteration count | Total CPU for 6 hour run (s) | Factor increase in 3d gridpoints over 12km | Factor CPU increase over 12km |
|---|---|---|---|---|---|---|
| **12km** | 146x182x32 | 300s | 22 | 23560 | - | - |
| **4km** | 190x190x38 | 60s | 6 | 119900 | 1.6 | 5.1 |
| **1km** | 300x300x76 | 30s | 35 | 1303900 | 8.0 | 55.3 |
| **1km (short tstep)** | 300x300x76 | 15s | 9 | 1900000 | 8.0 | 80.6 |

Table B2. Timing data for the models on the t3e supercomputer. For the work reported here the 1km model was run with a 30s timestep. The final row of this table shows data for the 1km model running with half this timestep is for comparison purposes only.

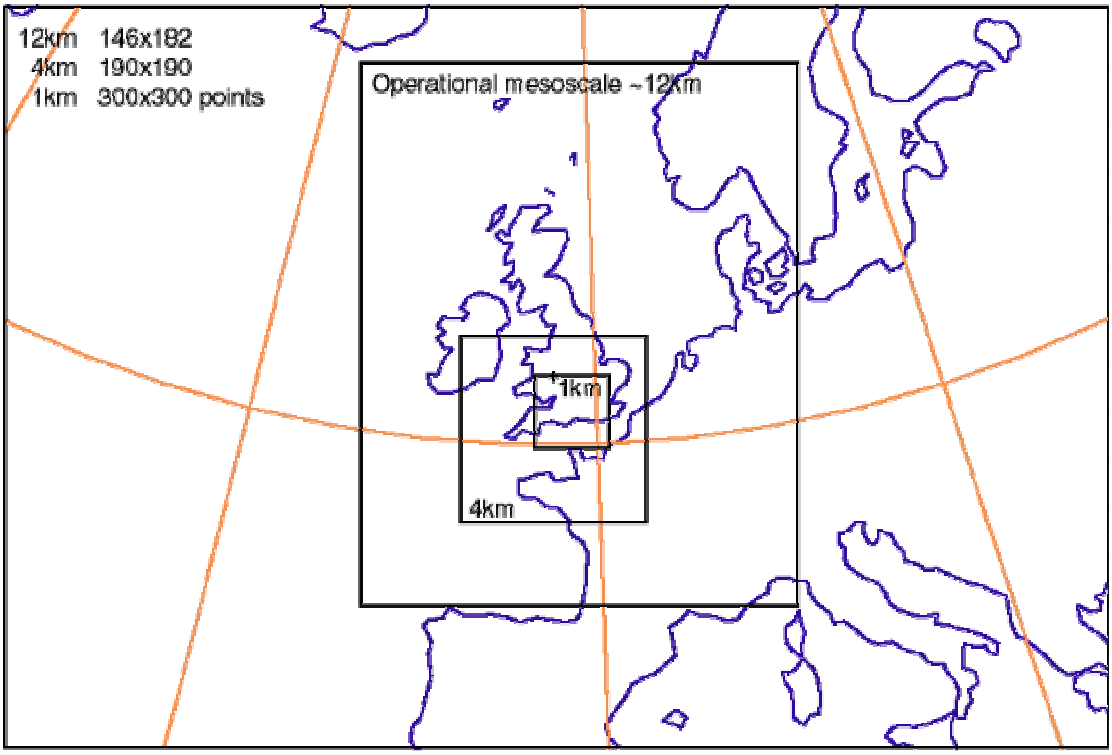

Figure 2.1 Domains used for the models

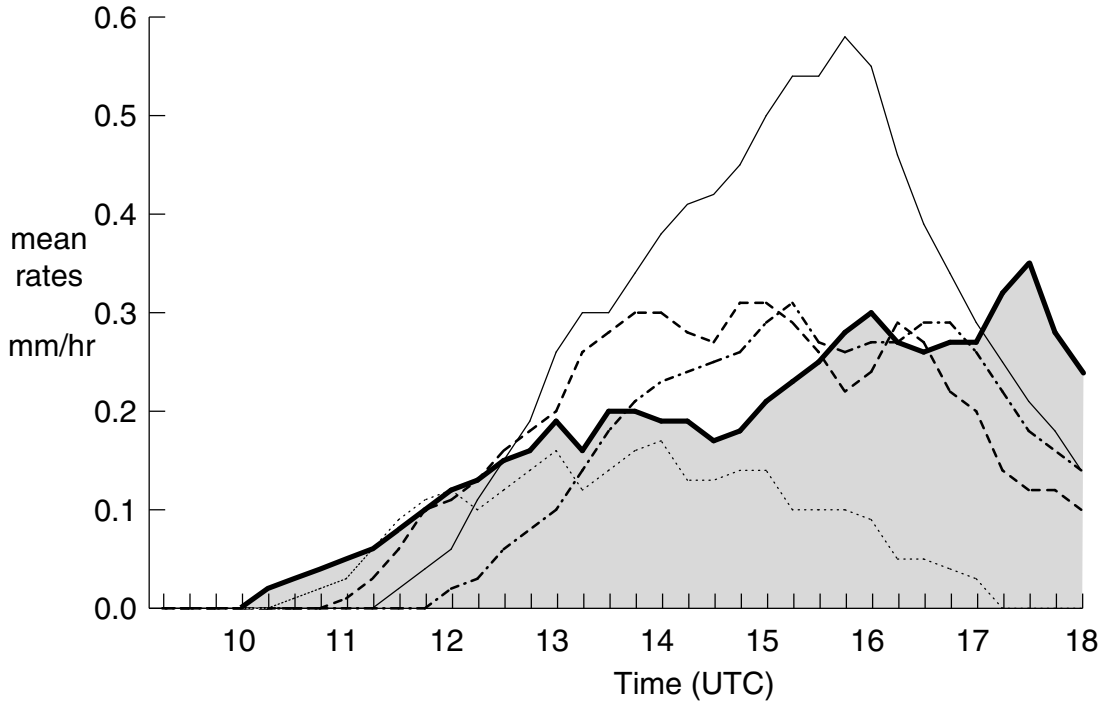


Figure 2.2 1km domain area averaged rainfall rates against time for 3rd May 2002. The thick solid line (shaded below) is the nimrod radar, the thin solid is the 4km model with no convection scheme, the dotted line is the 4km with the standard (i.e. fixed CAPE closure) convection scheme and the dashed-dotted is the CAPE dependent CAPE closure timescale scheme (c=0.1). Also shown as the dashed curve is the equivalent curve for the 1km model (without convection scheme).

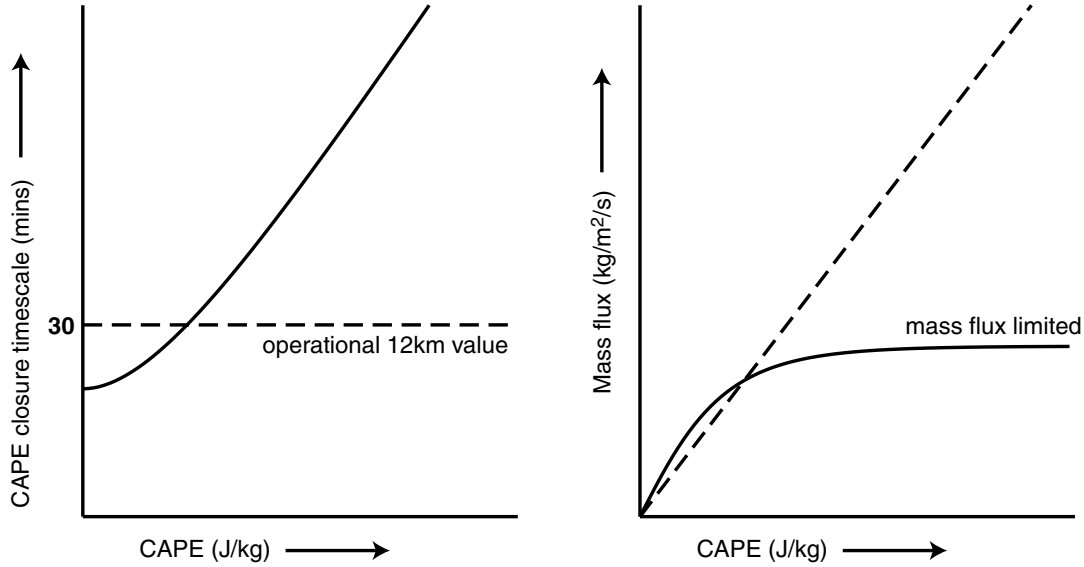


Figure 2.3. Graph of the function used to modify the CAPE closure timescale (t) in the convection scheme and the subsequent behaviour of the cloud-base mass flux (M) with CAPE.

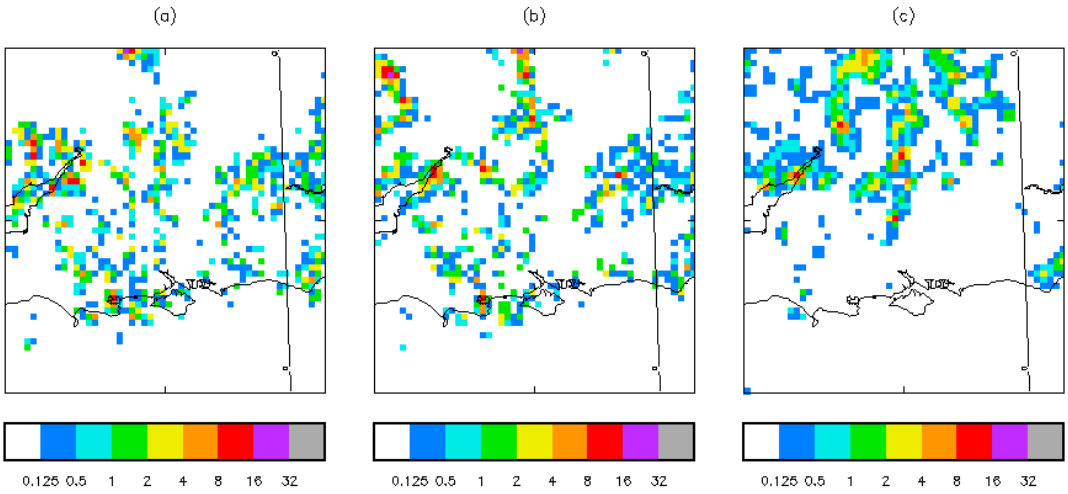


Figure 2.4 Rainfall rates for 13:45 UTC on 3rd May 2002. (a) shows 1km model without convection scheme but using boundaries from 4km model running with convection scheme, (b) shows same but with 4km model without convection scheme and (c) is radar data

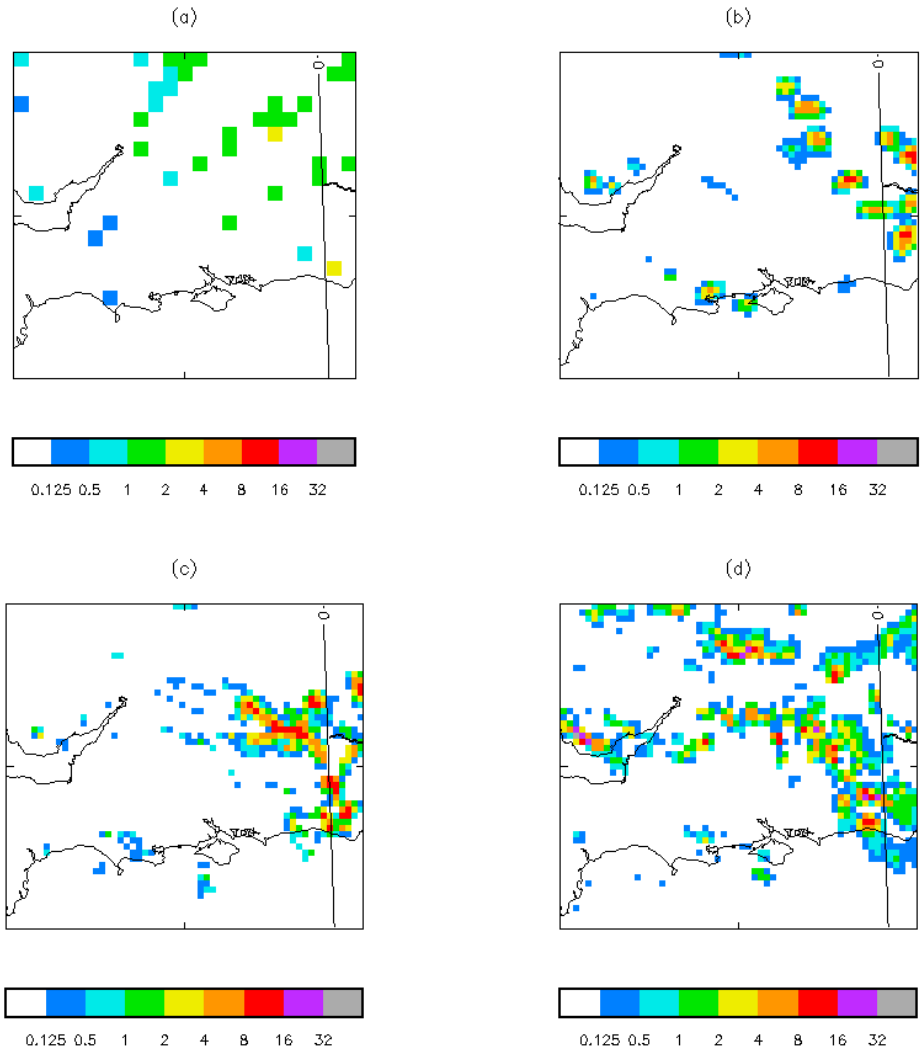


Figure 4.1 Precipitation rates at 15 UTC 13th May 2003 from the forecast initialised at 09 UTC for (a) the 12km model, (b) the 4km model and (c) the 1km model compared to the radar (d). The radar data is shown on a 5km grid and for straightforward comparison the 4km and 1km model data have been aggregated onto the same grid.

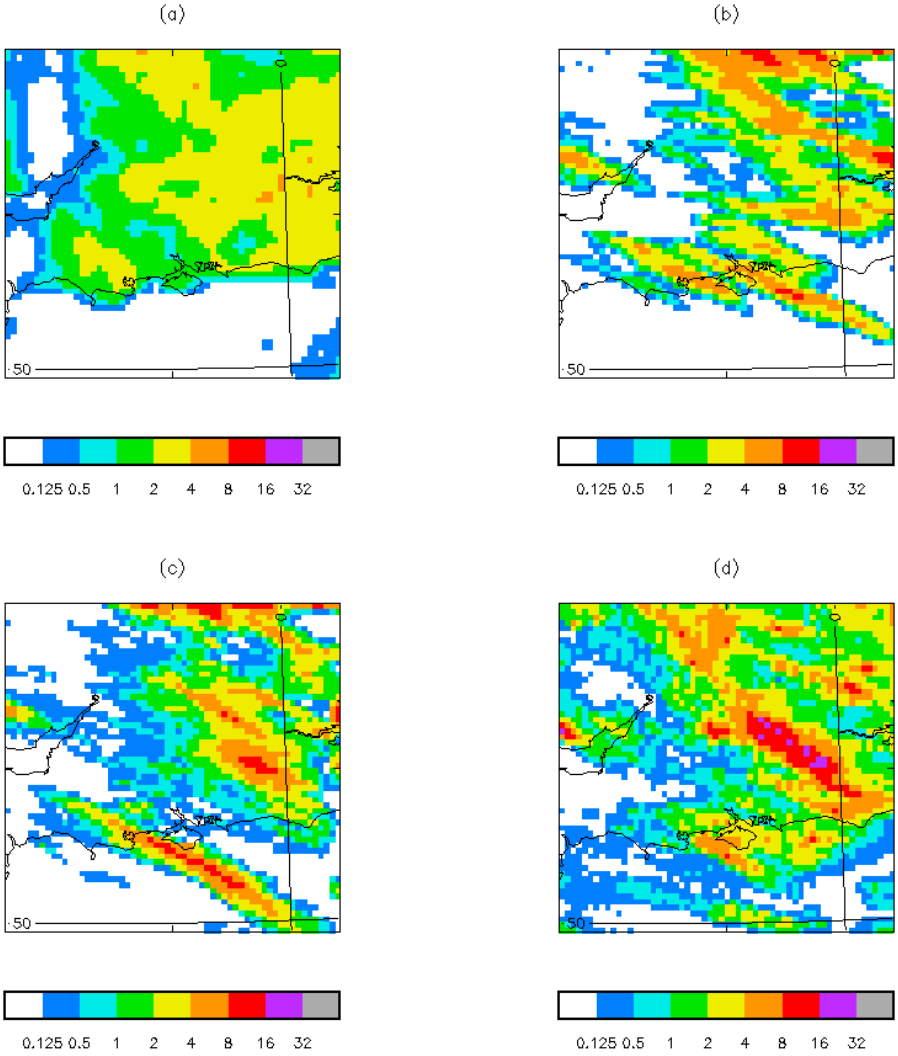


Figure 4.2 10-16 UTC rain accumulations for 13th May 2003. (a) is the 12km model, (b) the 4km model and (c) the 1km model and (d) is the radar data. As before the 4km and 1km data have been aggregated to the (5km) radar grid.

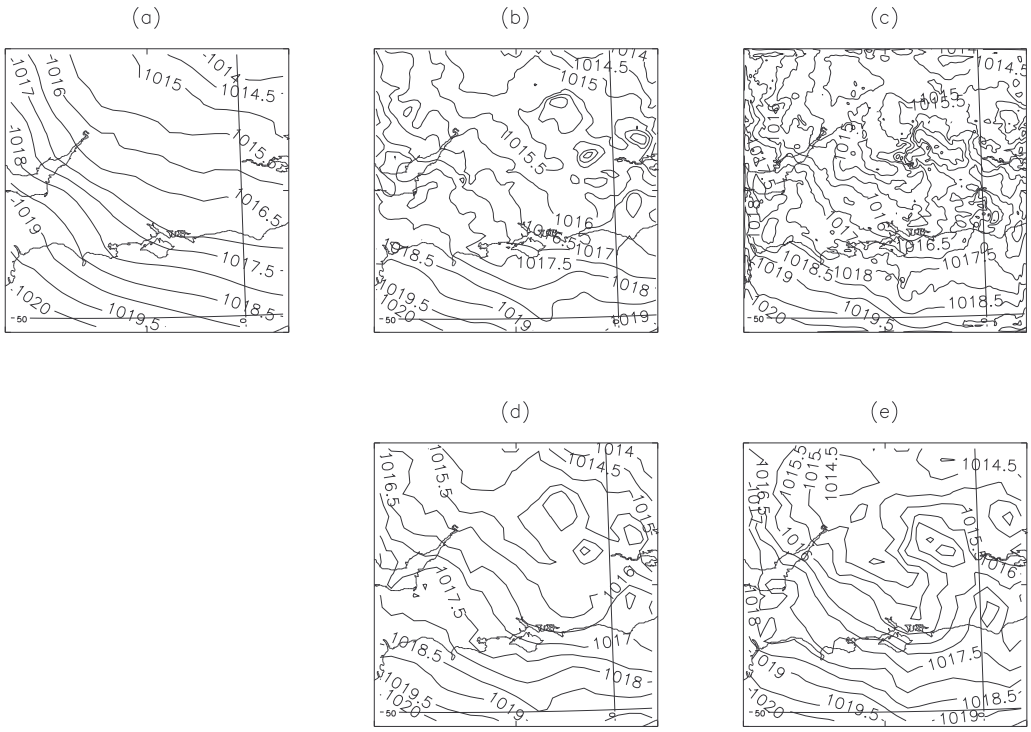


Figure 4.3 15 UTC mean sea level pressure for 13th May 2003 from the 09 UTC forecast. (a) is the 12km model data, (b) the 4km and (c) the 1km. (d) and (e) show the 4km and 1km fields but with the data aggregated onto a 12km grid. In all the plots the contours are at 0.5 hPa intervals.

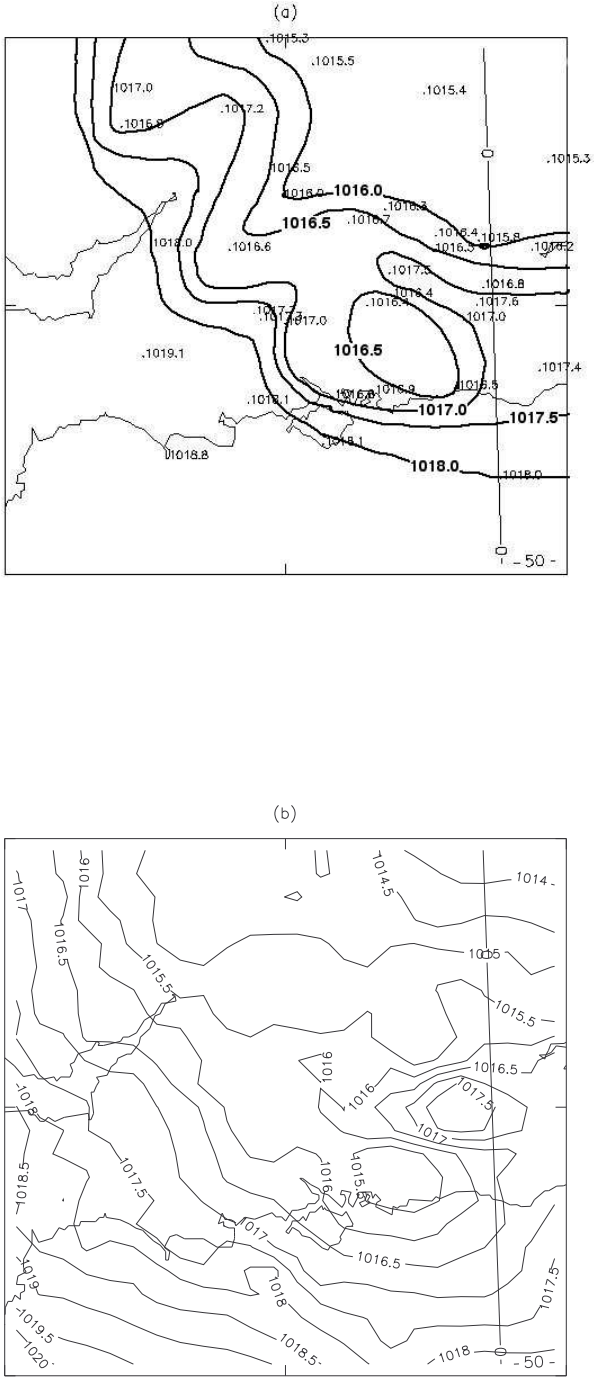


Figure 4.4. (a) Shows hand analysis for 16 UTC 13th May 2003 (bold lines and type) superimposed on observations. (b) shows corresponding 1km model field, aggregated to a 12km grid, with equivalent contour intervals.

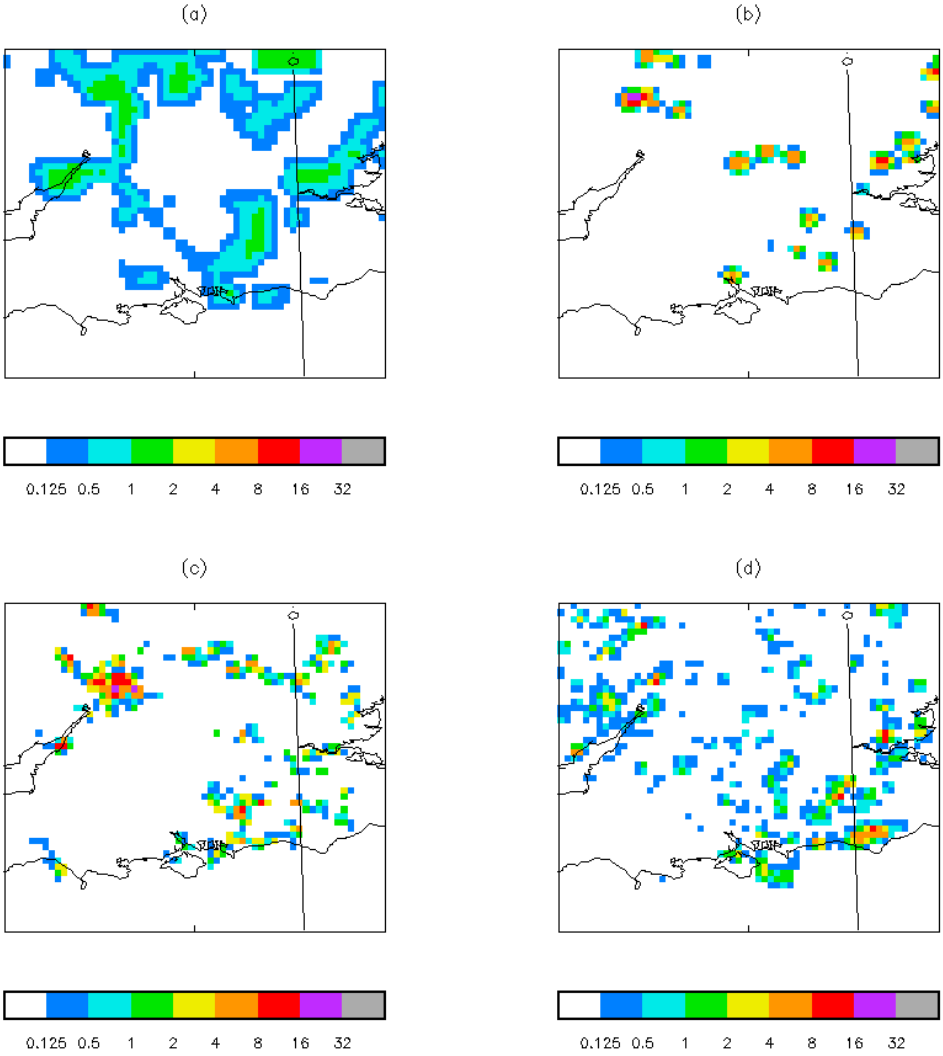


Figure 4.5 Precipitation rates aggregated to the 5km radar grid for 12 UTC 25th May 2003 from the forecast initialised at 06 UTC for (a) the 12km model, (b) the 4km model and (c) the 1km model compared to the radar (d).

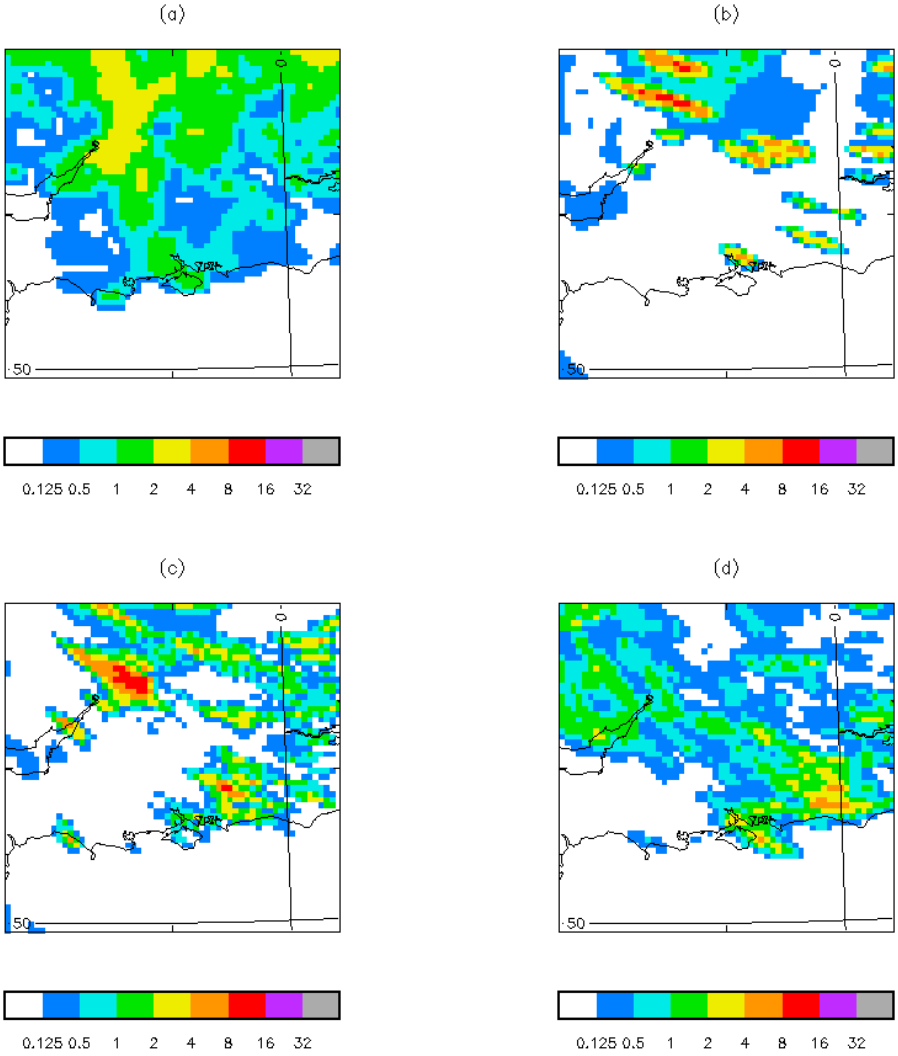


Figure 4.6 7-13 UTC rain accumulations for 25th May 2003. (a) is the 12km model, (b) the 4km model and (c) the 1km model and (d) is the radar data.

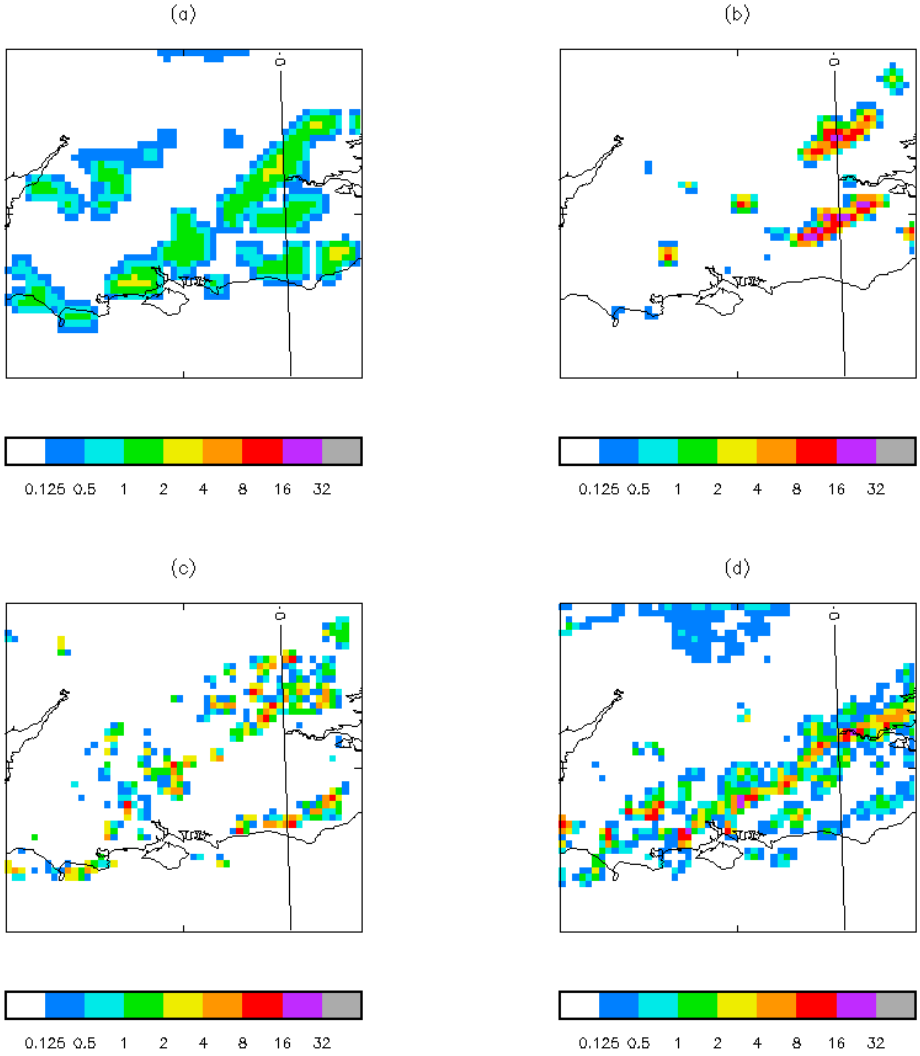


Figure 4.7 Precipitation rates aggregated to the 5km grid for 12:30 UTC 1st July 2003 from the 06 UTC forecast for (a) the 12km model, (b) the 4km model and (c) the 1km model compared to the radar (d).

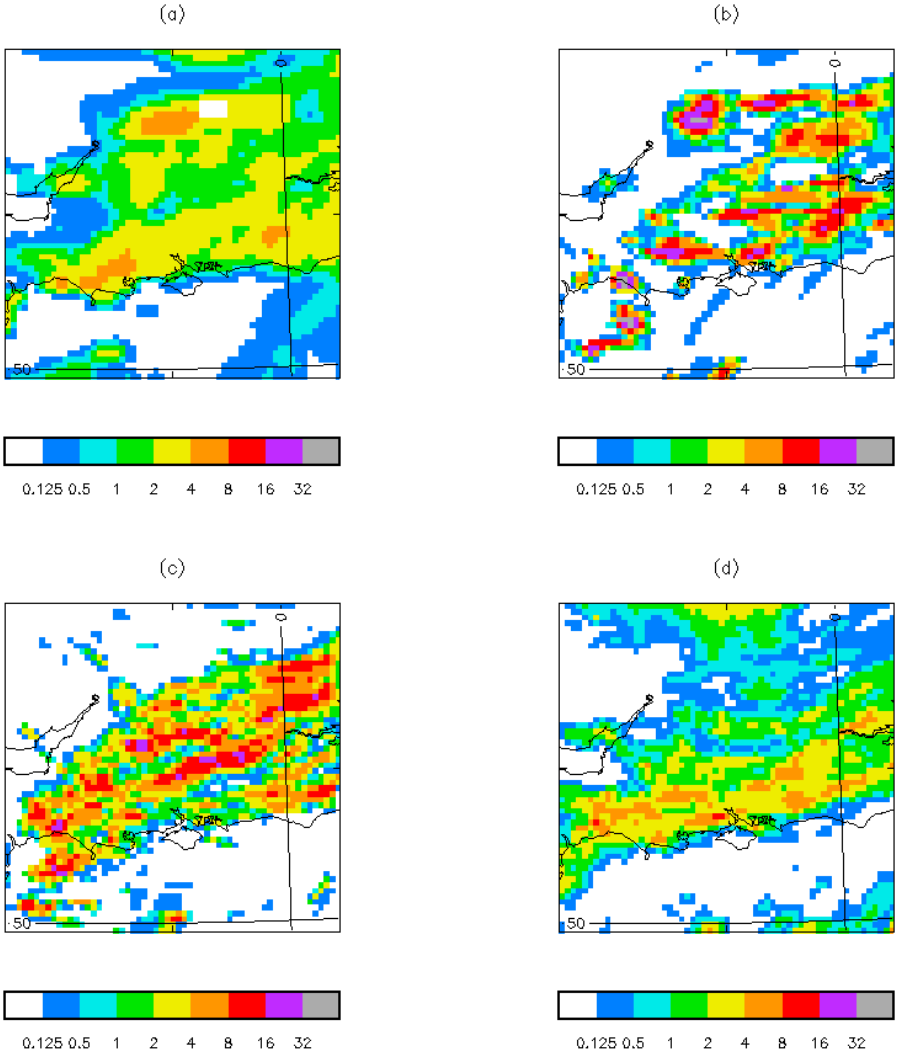


Figure 4.8 7-13 UTC rain accumulation aggregated to the 5km grid for 1st July 2003 from the 06 UTC forecast for (a) the 12km model, (b) the 4km model and (c) the 1km model compared to the radar (d).

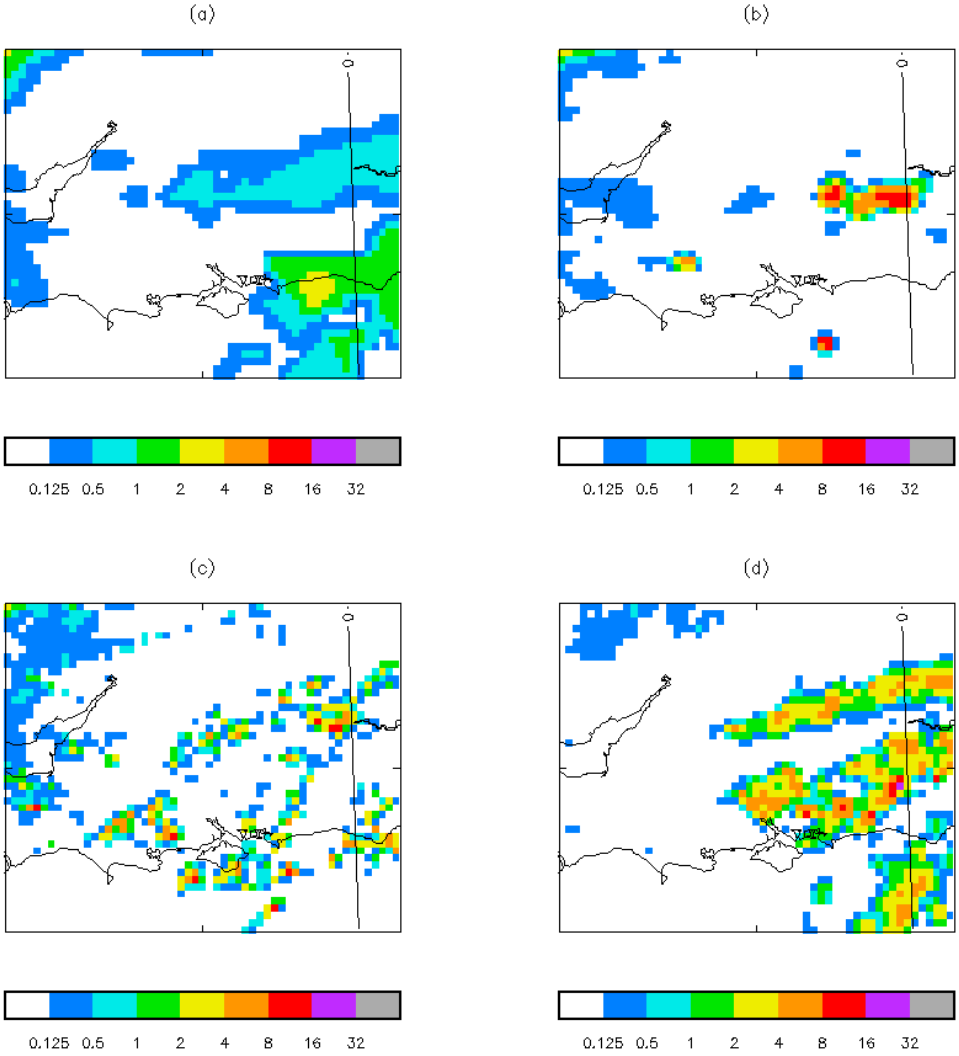


Figure 4.9 Precipitation rates aggregated to the 5km grid for 16 UTC 28th August 2003 from the 09 UTC forecast for (a) the 12km model, (b) the 4km model and (c) the 1km model compared to the radar (d). Note that in order to bring out the lower rainrates this figure only has a logarithmic scale.

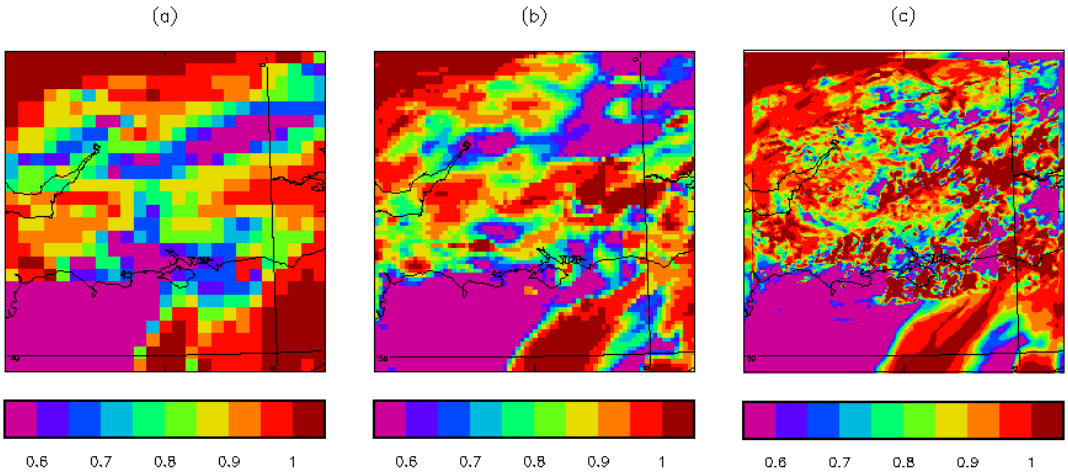


Figure 4.10 Total cloud fraction (max/random overlap) corresponding to model fields shown in fig 4.8. (a) is 12km model, (b) is 4km and (c) is 1km.

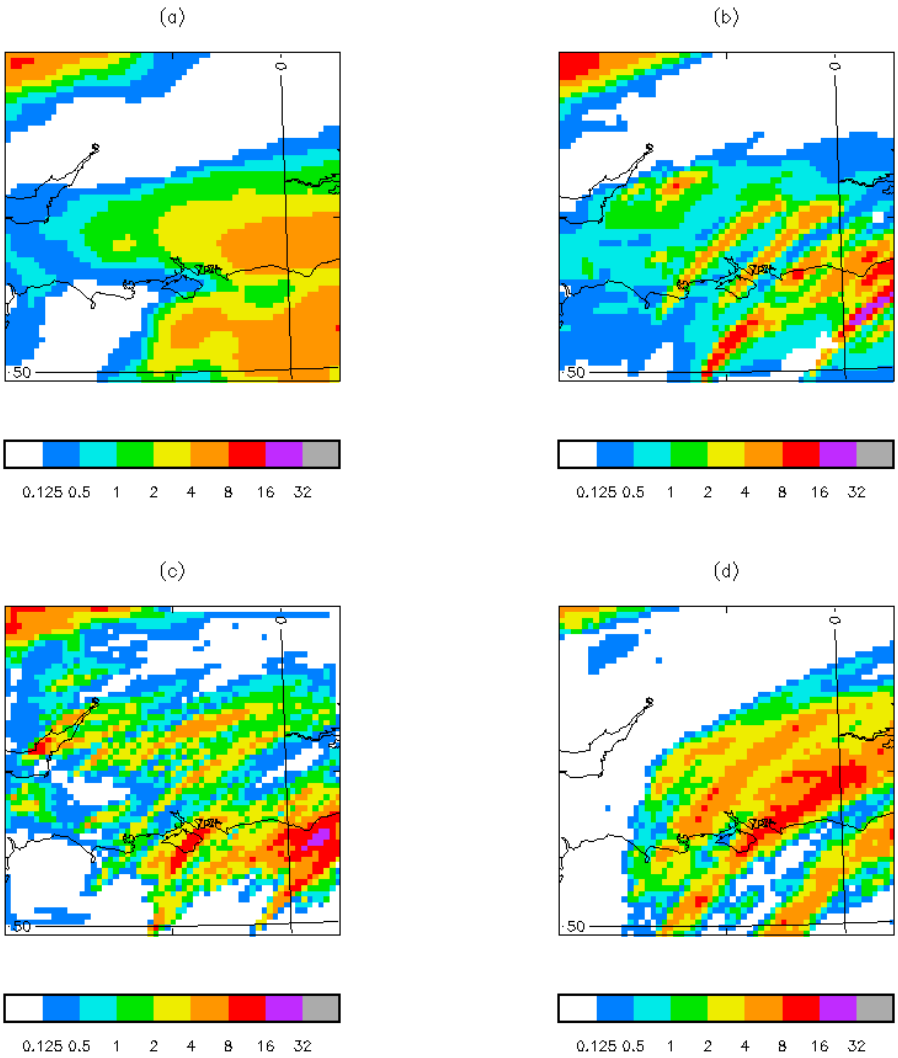


Figure 4.11 Rainfall accumulations from 10-16 UTC on 28th August 2003 from 09 UTC forecast aggregated onto 5km grid. (a) is 12km model, (b) 4km and (c) 1km. (d) is radar data.

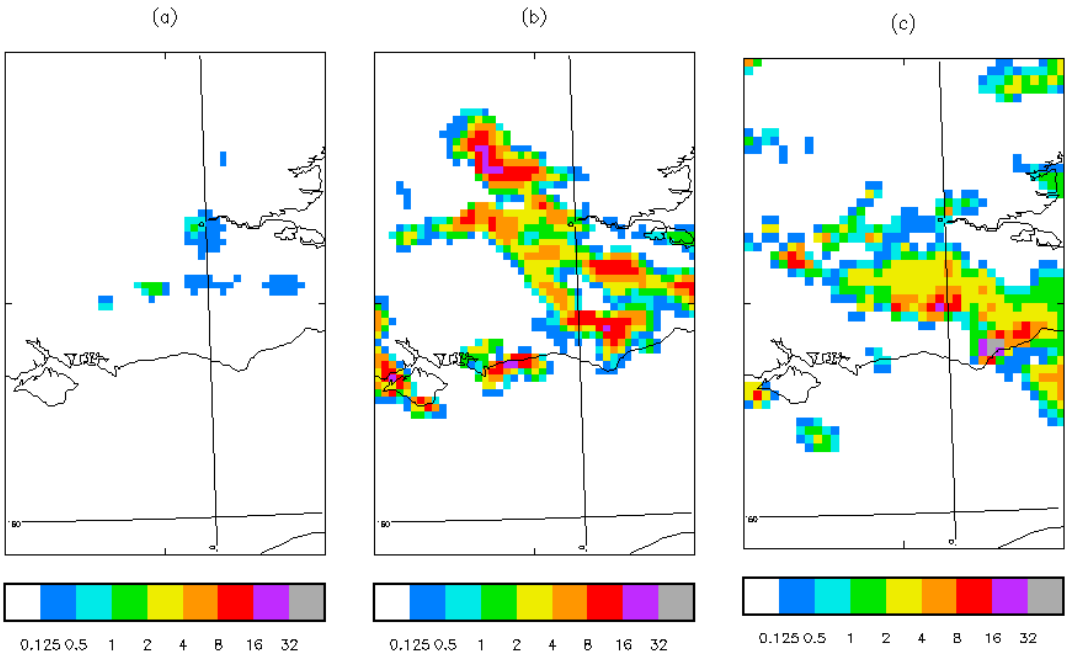


Figure 4.12. Precipitation rates for 17 UTC 13th May 2003. (a) is 4km forecast spinning up from 16 UTC T+1 12km analysis, (b) is same but 15 UTC forecast from continuous assimilation and (c) is the corresponding radar picture.

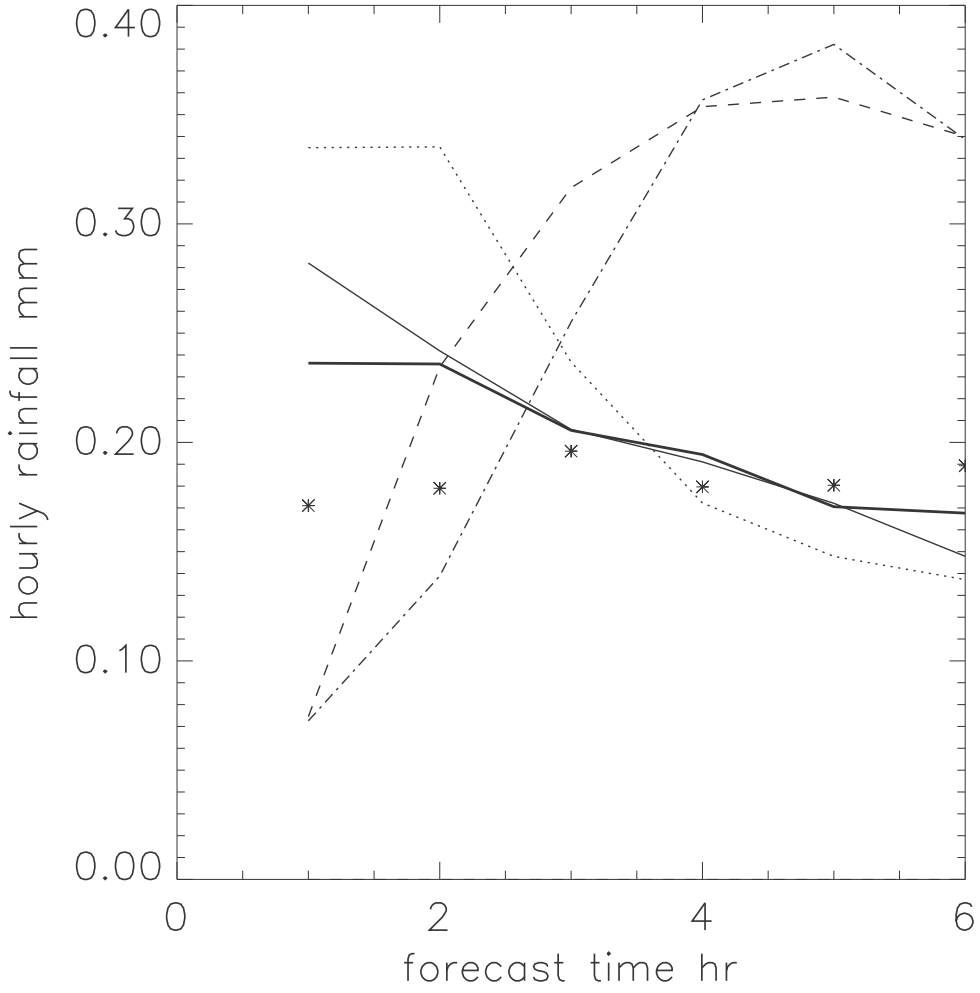


Figure 4.13 1km domain area averaged precipitation hourly accumulations as a function of forecast time averaged over all 16 forecasts used in this study (i.e. four forecasts for each case). Each point on the x axis represents an accumulation from the time to an hour later. The thick solid line is the 12km model, the dotted line the 4km model with assimilation and the thin solid line the 1km model with assimilation. The dashed line is the 1km model spinning up from each 12km analysis and the dashed dot line the equivalent 4km model. The stars represent the radar data.

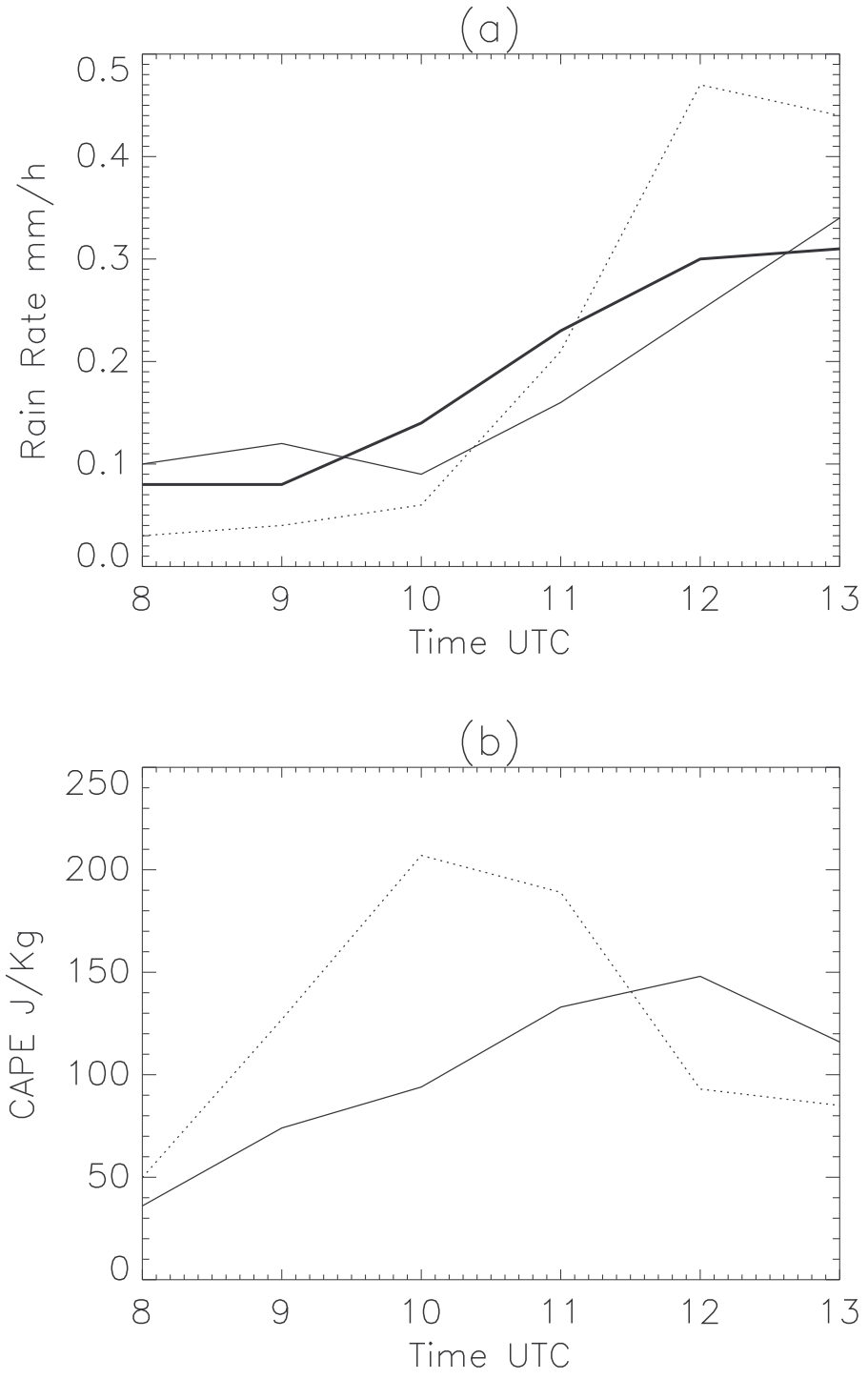


Figure 4.14 1km domain area averaged rain rates (frame a) and CAPE (frame b) plotted against time (UTC) for 06 UTC forecasts on 13th May 2003. Thin solid lines represent data from a run with assimilation and dotted lines a run spinning up from the 07 UTC T+1 12km analysis. The thick solid line in frame (a) represents the corresponding average rain rate from the radar data.

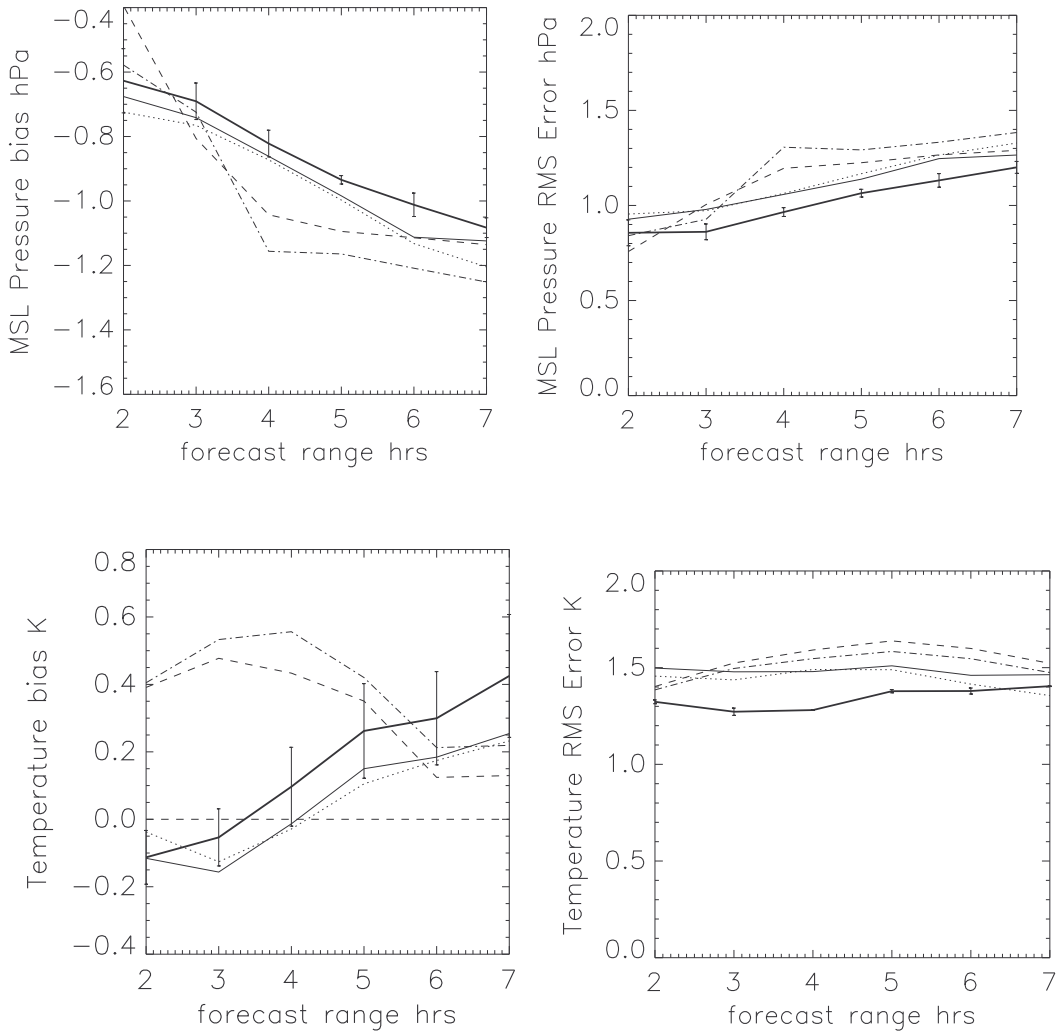


Figure 5.1 Bias and RMS errors between models and surface observations averaged over all forecasts of all four cases. The thick continuous line is the 12km model, the dotted line the 4km and the thin continuous 1km all with continuous assimilation. The dashed-dot line is the 4km model spinning up each cycle from 4km analyses and the dashed line in the 1km spinning up. The error bars indicate 1 standard deviation. For clarity this is shown for the 12km model only but this is representative of the spread in the other models also.

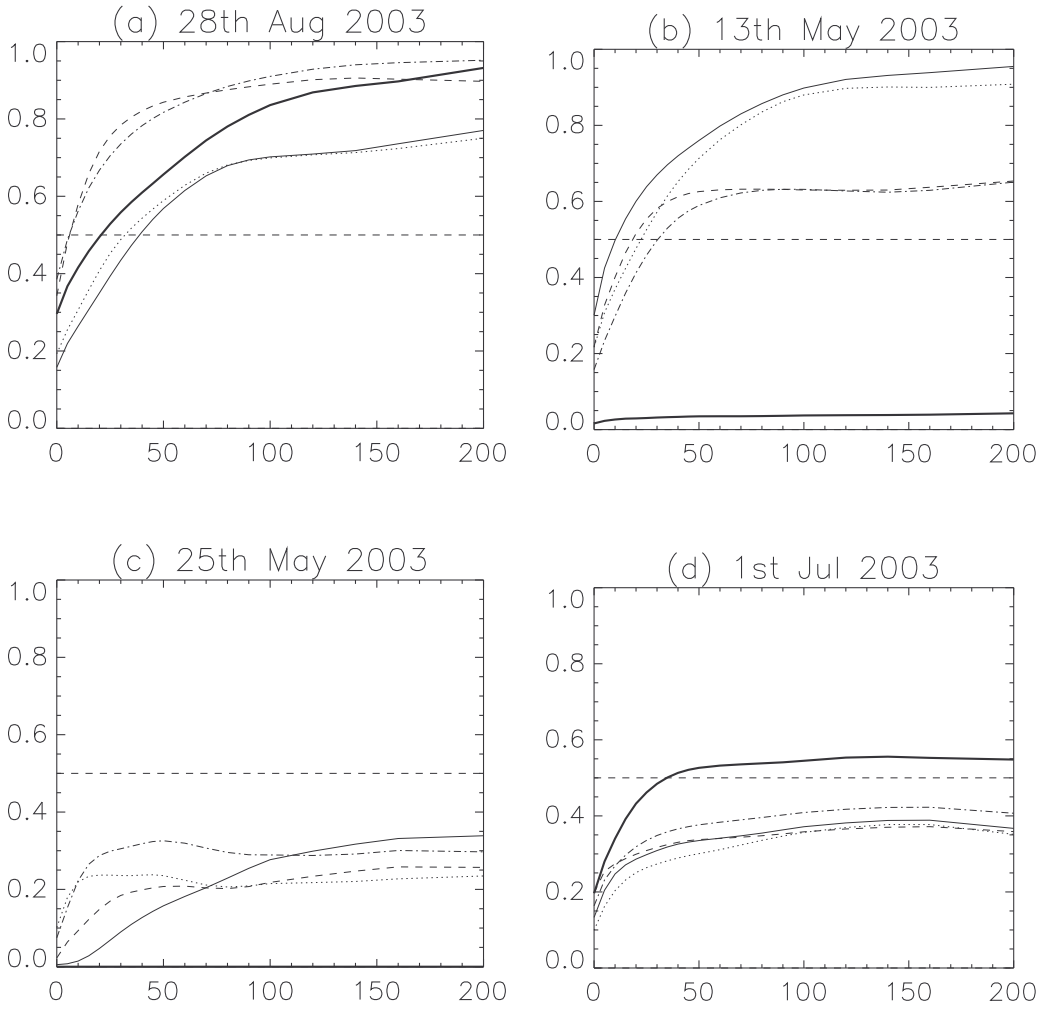


Figure 5.2 Fraction skill scores for a threshold of 4mm in 6 hours aggregated over the 4 forecasts of each case as a function of sampling radius (km). The thick solid line is for the 12km model (note that it runs along the x xis in frame (c) ). The dotted line is the 4km model with assimilation and the thin solid line the 1km model. The dash-dot line is the 4km model spinning up from the 12km analysis and the dashed line is the 1km model spinning up.

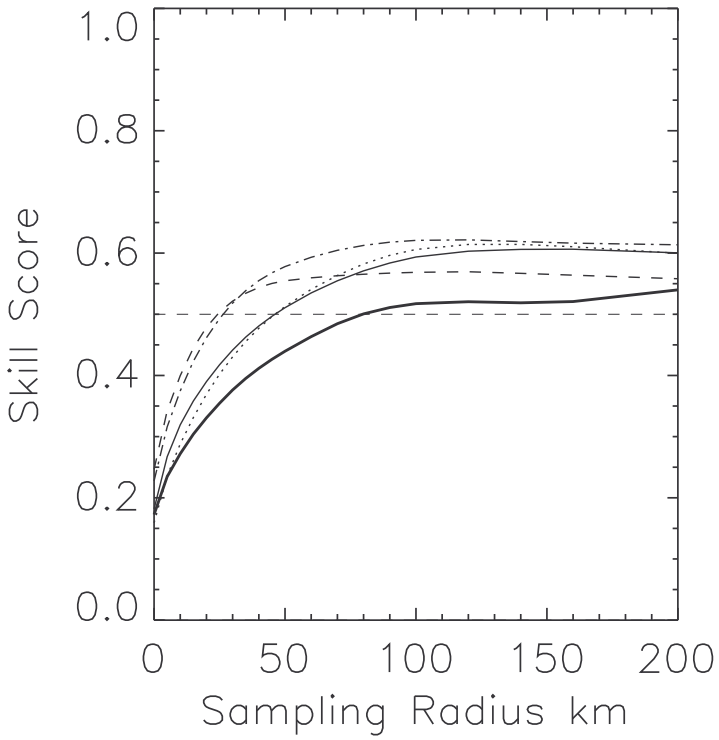


Figure 5.3 As figure 5.2 but including results from all four cases aggregated together.

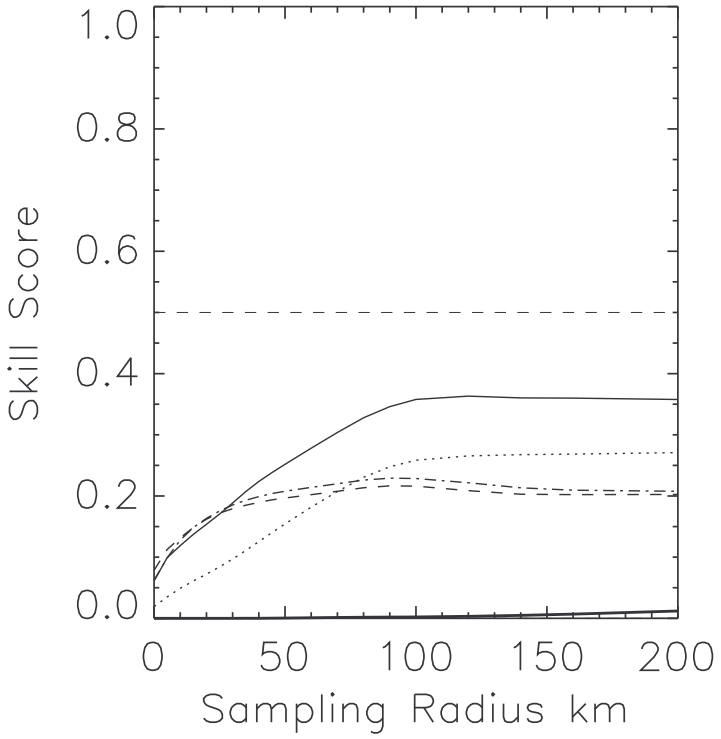


Figure 5.4 As figure 5.3 but for a rainfall threshold of 8.0mm in 6 hours.

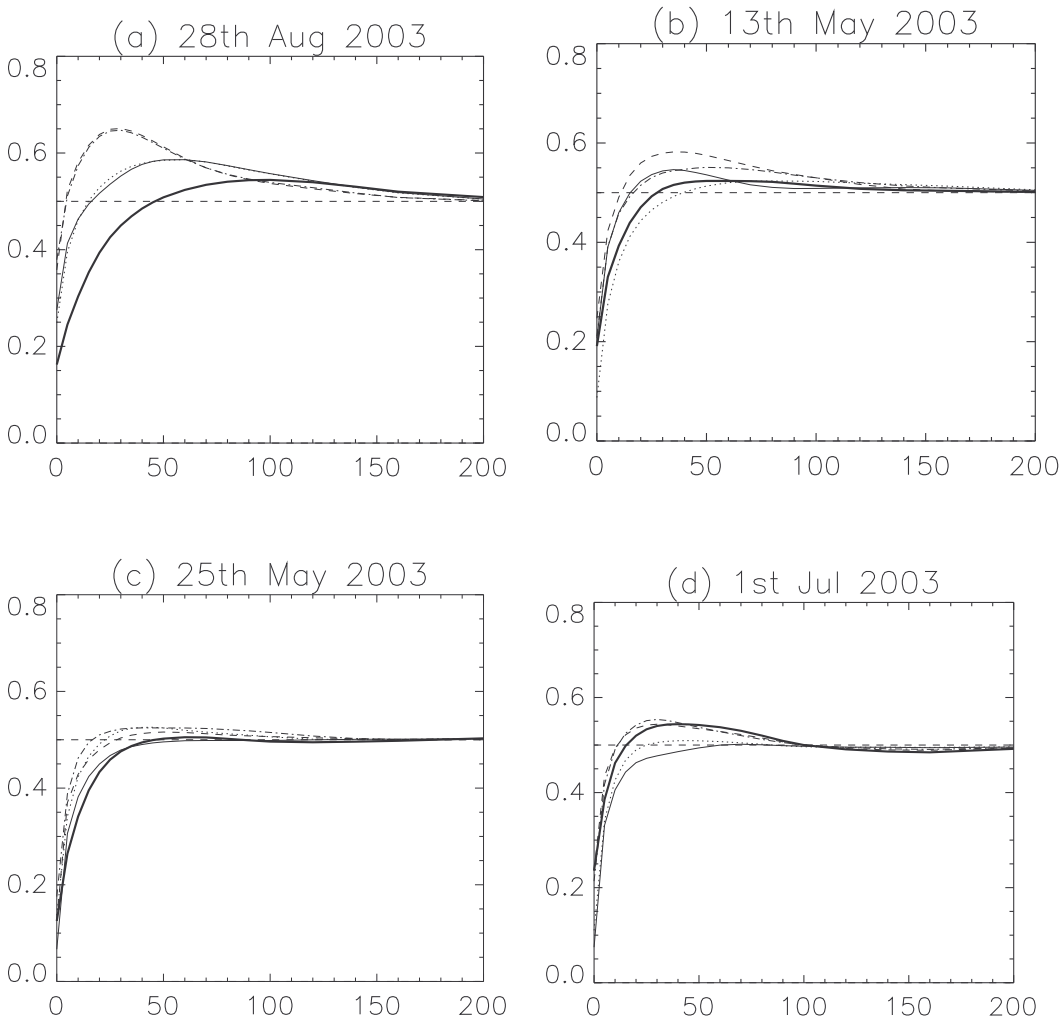


Figure 5.5 Brier Skill scores for top 10% threshold for each of the four cases.

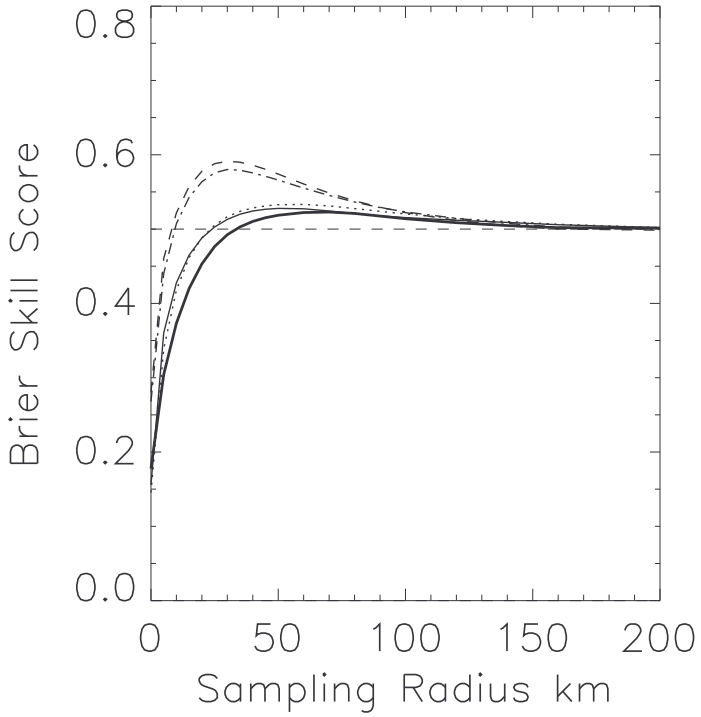


Figure 5.6 Brier Skill scores for top 10% threshold for the four cases aggregated together.

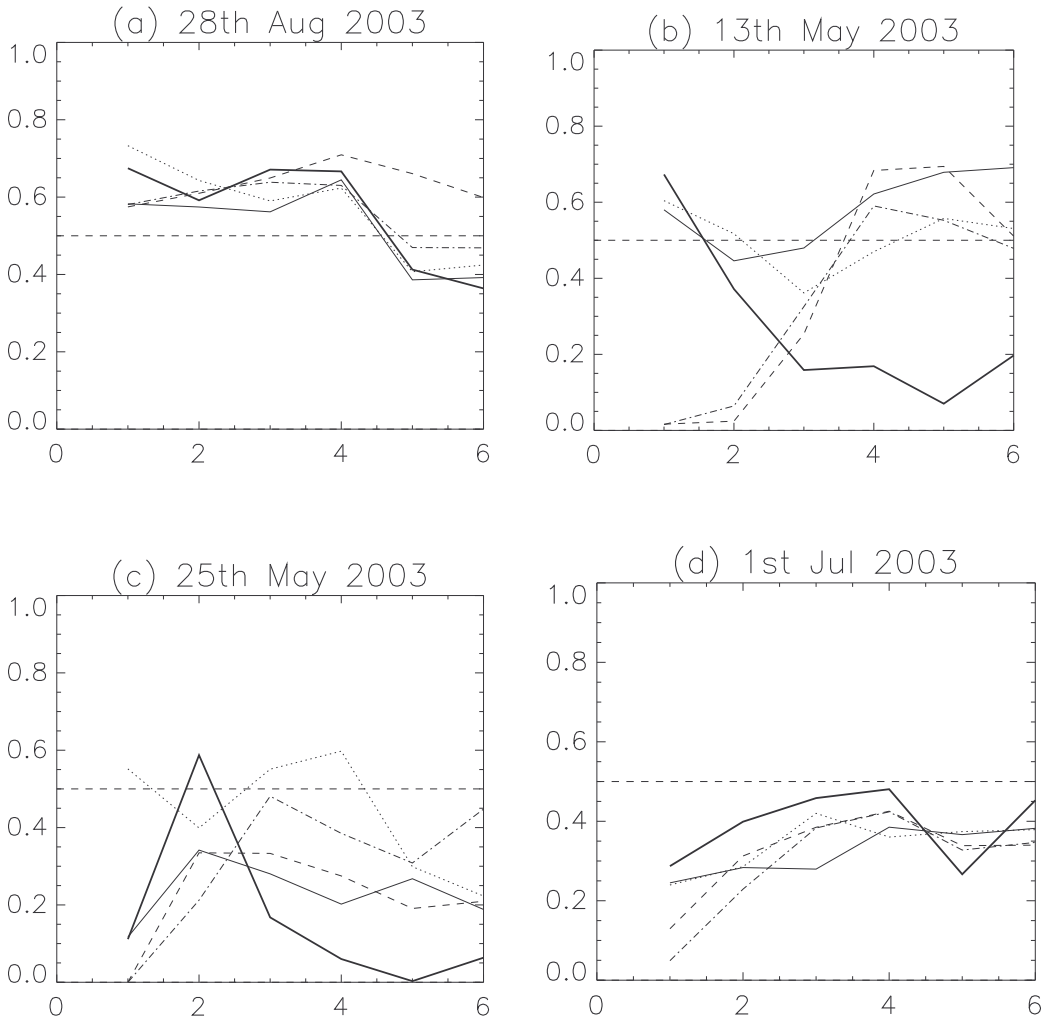


Figure 5.7 One hour accumulation fraction skill scores for a threshold of 1.0mm plotted against forecast range for a fixed sampling length of 55km. The four frames show the results from each of the four cases separately. The lines have the same meaning as in figure 5.2

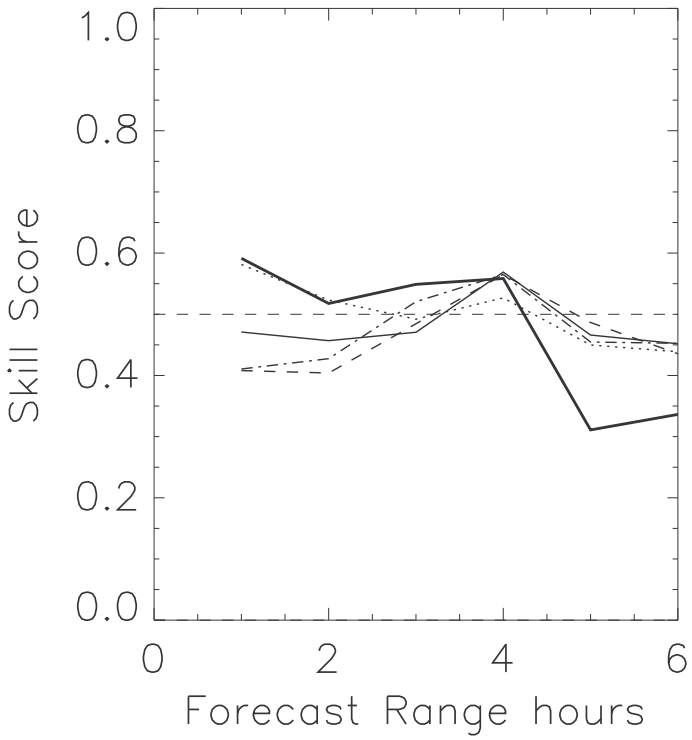


Figure 5.8 as figure 5.7 but aggregating the results from all four cases together.